Growth of graded quantum wells for THz polaritonics

by

Chris Deimert

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2021

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:	Gottfried Strasser, Professor, Electrical Engineering and Information Technology, Vienna University of Technology (TU Wien)
Supervisor:	Zbigniew R. Wasilewski, Professor, Electrical and Computer Engineering, University of Waterloo
Internal Member:	Dayan Ban, Professor, Electrical and Computer Engineering, University of Waterloo
Internal Member:	Na Young Kim, Associate Professor, Electrical and Computer Engineering, University of Waterloo
Internal-External Member:	Robert Hill, Associate Professor, Dept. of Physics and Astronomy, University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Rectangular quantum wells have long dominated the landscape of layered nanostructures. They exhibit a rich variety of physics and can be reliably grown with techniques such as molecular beam epitaxy. Rectangular wells represent only a fraction of the possible design space, however: much less explored have been alternative structures with continuously varying potential profiles. This is not for want of applications. Parabolic quantum wells (PQWs), wells with a quadratically varying profile, have been recently identified as a potential key ingredient for terahertz (THz) polaritonic devices. The development of THz devices remains an important challenge with myriad applications that are only just beginning to be unlocked. We aim toward this important challenge by developing THz polaritonics based on PQW active regions, bringing to bear the fascinating physics of polariton quasiparticles – part light, part matter – and one of the most ubiquitous elements in physics, the harmonic oscillator.

Using non-square quantum wells comes with challenges, however. Growing these structures with molecular beam epitaxy requires time-varying flux from the effusion cells, which is difficult to produce reliably due to the slow thermal response of the cells. While this problem can sometimes be bypassed through the use of digital alloys, there are limits on the quality of material that can be produced in this way. We develop an approach for growing smoothly graded quantum wells with molecular beam epitaxy, compensating for the effusion cell's thermal behaviour via a simple linear dynamical model. We further provide an iterative scheme to correct for any lingering errors. With this technique, we demonstrate the ability to grow smoothly graded $Al_x Ga_{1-x} As$ PQWs for THz polaritonics, with a root-mean-square Al composition (x) error of just ± 0.0018. This accuracy is achieved at the standard growth rates (0.15–0.25 nm/s) necessary for thick structures of several micrometres or more. The approach is quite generally applicable beyond $Al_x Ga_{1-x} As$ PQWs, and could be used for other materials or composition profiles, or for other situations where precise time-dependent flux control is required.

We further study the properties of continuously graded $Al_x Ga_{1-x} As$ PQWs grown in this manner, both theoretically and experimentally. Theoretically, we pull together existing work in the literature to build a numerical semiclassical model of quantum well absorption, which includes the multi-subband plasmonic interactions that appear at the doping levels required for THz polaritonics. This, combined with electromagnetic modelling allows us to study the design aspects of THz quantum wells relevant to polaritonic devices. We explore the possibility of studies on polariton-polariton scattering, and examine the quality of quantum well active region necessary for such studies. We further explore the possibility of new active region designs incorporating half parabolic quantum wells.

Experimentally, we probe PQWs grown with the continuous grading method, using reflection measurements in a metal-insulator-metal cavity and absorption measurements in a multipass geometry. We demonstrate robust oscillation at 2.1 THz and 3 THz for structures containing 8, 18, and 54 PQWs. Absorption at room temperature is achieved, which is as expected from a parabolic potential but would typically be impossible with square quantum wells in the THz. Linewidths below 12 % of the central frequency are obtained up to 150 K in two of the samples, with a very small 3.9 % linewidth obtained in one sample at 4 K. Furthermore, we show that the system correctly displays an absence of nonlinearity despite electron-electron interactions – analogous to the Kohn theorem. The high quality of these structures already opens up several new experimental vistas.

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Zbig Wasilewski. Your support, guidance, and experience have been invaluable during my time in Waterloo. I am inspired by your curiosity and passion for science, and I have always been struck by your ability to remain calm in the face of challenges and setbacks. I truly hope that I can carry some piece of that forward.

I would like to thank Profs. Gottfried Strasser, Rob Hill, Na Young Kim and Dayan Ban for the time they've dedicated to serving as my committee members. I would especially like to thank Prof. Kim and her post-doc Dr. Taehyun Yoon for their work on the photoluminescence excitation measurements in this thesis. Thanks to both Prof. Kim and Prof. Ban for many good discussions and collaborations over the years.

I would like to thank Dr. Wojtek Pasek for his tireless efforts in plasmon modelling, and for his patience in guiding us experimentalists through the dark jungles of $\mathbf{k} \cdot \mathbf{p}$ theory.

I would like to thank the folks at the Centre for Nanoscience and Nanotechnology in Paris – particularly Dr. Raffaele Colombelli, Dr. Jean-Michel Manceau and soon-to-be-Dr. Paul Goulain – who were kind enough to host me there for several trips and who were heavily involved in the many of the designs and experiments presented throughout this thesis. I certainly could not have done this on my own, and it was a pleasure to work with all of you.

I should also thank the folks at École normale supérieure, with particular thanks to Jacques Hawecker who not only helped us with time domain spectroscopy measurements, but who kindly lent me his guitar while I stayed in Paris.

I feel I must give special thanks to my dog, Sterling, who has frequently encouraged me to take breaks from writing my thesis to play and go on walks. She is always selflessly looking out for my mental well-being.

More seriously though, I would like to thank my parents, who always encouraged my sense of curiosity. And I would like to thank all of my colleagues and friends and family who have supported me both in my work and in my life outside of research. I started listing your names, but there are just far too many of you. Believe me that each one of you has made a lasting impact. Doing a PhD is no easy task, and life has a way of surprising us with difficulties and trials. Many of you have helped me through the difficult times: probably more than you realize. This – my whole life in the last five and a half years in Waterloo – really wouldn't have been possible without you.

Table of Contents

List of Figures

List of Abbreviations

2DEG two-dimensional electron gas 53

BFM beam flux-monitoring ion gauge xi, xii, 57–59, 61–63, 76–78, 98, 110, 111, 113, 114

CBE chemical beam epitaxy 65

FTIR Fourier transform infrared xiv, 82, 89, 90

FWHM full width at half maximum 48

HPQW half parabolic quantum well xiv, xv, 2, 92–99, 101

HR-XRD high-resolution x-ray diffractometry xii, 59, 76, 77

IR infrared 1, 8–10, 18, 20, 24, 28, 31, 43, 50, 65

ISB intersubband x, 12, 18, 20–22, 24–26, 28, 30, 32, 33, 38, 40–44, 48, 49, 51, 54, 80, 82, 84, 85, 88, 91

MBE molecular beam epitaxy ix, xi, xv, 1, 2, 9, 23, 53–61, 63–67, 69, 79, 80, 91, 97, 98, 100, 101, 110–113

MIM metal-insulator-metal ix, x, xiv, 1, 2, 23, 24, 26–29, 31–33, 45, 47, 80, 88, 89, 100, 103

MOCVD metal-organic vapour-phase epitaxy 65

MSP multisubband plasmon 40–43, 48, 50, 51, 100

PBN pyrolytic boron nitride 69

PID proportional, integral, derivative xi, 55, 57, 66, 69, 71, 73

PL photoluminescence xiii, 87

PLE photoluminescence excitation xiii, 2, 86–88

PQW parabolic quantum well iv, x–xiv, 1, 2, 34–38, 43, 49–52, 54, 66, 72, 73, 76–92, 94, 96, 98, 100, 101

QCL quantum cascade laser 8–10, 28, 29, 54

QW quantum well x, xi, xiv, 1, 2, 14, 20–22, 24–26, 28, 31–37, 39–41, 43–53, 68, 84–86, 89, 92, 94–96, 98, 100, 101, 103–109, 122, 123, 125

RCWA rigorous coupled-wave analysis 24

STM scanning tunnelling microscopy 5

TDS time-domain spectroscopy 5

TE transverse electric 104

THz terahertz iv, ix, 1–11, 18–25, 28–33, 35, 38, 40, 43, 47–51, 53, 54, 65, 66, 73, 79, 80, 82, 84, 90–92, 94, 100, 101, 108

TM transverse magnetic xiv, 24, 45, 89, 104

UHV ultra-high vacuum xi, 54, 56–58

Chapter 1

Overview

As the title suggests, in this thesis we will explore the growth of continuously graded quantum wells in the context of terahertz (THz) polariton devices. We will study parabolic and half-parabolic quantum wells (QWs) from several perspectives, including theoretical modelling, molecular beam epitaxy (MBE) growth, and characterization through absorption and reflection spectroscopy. Overall, our goal will be to gain a better understanding of the active regions used in THz polaritonics, and to develop techniques for improving their performance.

In the first section of the thesis, we will begin with motivation and background. Ch. 2 will explore why it is still important to develop THz photonic devices, surveying some potential applications and remaining challenges. Ch. 3 will then lay the foundation for polaritonic devices in general by looking at exciton-polariton lasers, which have received much attention in the optical and infrared (IR) frequencies.

After our brief background in THz and polaritons, we will combine the two in Ch. 4 to explore the design of THz polariton devices. We will see some of the challenges on the road towards THz polaritonics, and we will examine dispersive metal-insulator-metal (MIM) cavities (Chs. 4.2 and 4.3) as a potential platform on which to address some of these challenges – particularly those related to polariton scattering. In Ch. 4.3 we will see that, for such studies, the quality of the QW active region turns out to be a key ingredient. I should note here that the conceptual work in Ch. 4 was done in collaboration with Dr. Raffaele Colombelli's group at the University of Paris-Saclay, employing the Matlab code they developed for simulating MIM cavities.

In Ch. 5, we will take a closer look at modelling QW active regions. At the doping levels required for polariton devices, modelling these wells is a non-trivial matter involving interactions between several intersubband transitions, which can give rise to so-called multisubband plasmon modes. While much of the theory of these modes has been established already in recent years, there are no "out of the box" numerical codes available. Thus, to design the structures studied in this thesis, it has been critical to implement these theories in software. We have done this by piecing together existing software and models from the literature, additionally offering some clarifications and small extensions which are relevant to the particular problems studied herein. We explore both quantum (Ch. 5.3) and semiclassical models (Ch. 5.4) for multisubband plasmons. We also point to some weaknesses in the existing theories which will need to be addressed moving forward. Some of the modelling effort in Ch. 5 was born of discussions with Dr. Wojciech Pasek, and he has done significant work (not presented in this thesis) in the direction of addressing the aforementioned weaknesses.

Finally, once the theory is developed, we end Ch. 5 by studying several QW designs for THz polariton active regions (Ch. 5.7). We explore the behaviour of square QWs and parabolic quantum wells (PQWs)

including the effects of plasmon-plasmon interactions at high doping, and we see why PQWs are preferable to square QWs for THz polaritonics. This, combined with observation of Ch. 4.3 that small linewidths are important for scattering-based devices, motivates us to turn our attention to the growth of high-quality PQWs.

Our study of PQW growth begins with some fundamental aspects of MBE in Ch. 6, especially regarding flux calibrations. While these techniques may not be new, they have not always been consistently applied throughout the MBE community, and are worth examining more closely. We will take the opportunity to clarify and justify certain approximations that are often taken for granted.

In Ch. 7 we develop a new MBE technique for growing continuously graded alloys, which we will eventually use to grow PQWs. Graded alloys are not new in MBE, but it has remained difficult to obtain very precise control over time-varying cell fluxes – particularly if one is trying to grow at the fast growth rates required for thick structures. We develop (Chs. 7.1–7.5) and test (Chs. 7.6, 7.8 ,and 7.9) an approach for modelling effusion cell dynamics, which can be used to achieve precise compositional control at standard growth rates. We also provide an iterative correction scheme (Ch. 7.7) which can be used to compensate for weaknesses in our effusion cell model if extra precision is required. Although we focus on varying Al fluxes for $Al_x Ga_{1-x} As$ growths, the technique should be quite generally applicable.

In Ch. 8, we use our MBE technique to grow THz PQWs, designed for use in polaritonic devices. We examine their absorption properties, knowing as we do from Ch. 4.3 that it is important to have active regions with small absorption linewidths. We demonstrate (Chs. 8.1 and 8.2) continuously graded THz PQWs with very small linewidths, as low as 3.9 % of the absorption centre frequency. I should note that the efforts in this chapter were done in close collaboration with Dr. Colombelli's group, who performed many of the absorption and reflection measurements, as well as the MIM cavity fabrications. We also make use of photoluminescence excitation (PLE) measurements (Ch. 8.3) that were performed by Dr. Na Young Kim and Dr. Taehyun Yoon, and were analyzed by Dr. Wojciech Pasek. This demonstration in Ch. 8 of high-quality absorption in continuously graded PQWs already represents a step forward in active regions for THz polaritonics. At minimum, this should simplify experimental setups by loosening the requirements for low temperature and eliminating the need for liquid helium.

Finally, in Ch. 9, we look forward towards possible alternatives to PQWs, which would allow for more flexibility in the active region design for THz devices. In particular, we examine whether half parabolic quantum wells (HPQWs) can provide high-quality THz absorption resonances. Using the models of Ch. 5, we find (Ch. 9.1) that HPQWs in their basic form are not expected to perform well at room temperature, particularly with the doping levels required for polaritonics. However, we show (Chs. 9.2 and 9.3) that good performance should be recoverable if the composition profile is modified to a slightly shifted parabola. We grow these shifted HPQWs samples using our continuous alloy growth technique and demonstrate preliminary absorption results (Ch. 9.4). Work is still needed in this regard, but HPQWs with a correction to the composition profile could form the basis for interesting new active regions in THz polaritonics.

Chapter 2

Background: a short survey of THz devices

Let us begin with some background on terahertz (THz) devices. The THz range of the spectrum ($\sim$ 0.3–10 THz) has been historically underdeveloped. These frequencies are particularly challenging, being too high for standard microwave engineering techniques and too low to be addressed by the semiconductor bandgap engineering used in standard optical devices. Because of this, while microwave and optical sources enjoyed enormous growth through the 20$^{\text{th}}$ century, THz sources tended to remain bulky and expensive. Over the last few decades, there has been an explosion of development, with an eye towards numerous applications that arise as THz devices become smaller and cheaper.

In this chapter, we survey the landscape of THz devices, looking at some of their potential applications and current status. The goal here is not to perform an exhaustive review of THz devices, but simply to provide some context and motivation for the later discussion of strong-coupled devices in the THz.

2.1 Applications of THz devices

The recent interest in THz devices has been largely driven by the promise of practical applications in a wide range of sectors. Many review articles and books have already discussed these applications in detail, e.g., Refs. [1–4]. Here we will take a brief look at three classes of applications: (1) THz communications, (2) quantum information, and (3) THz imaging and spectroscopy.

2.1.1 THz communications

See, e.g., Refs. [1, 5] for further discussion.

Wireless communications are an integral part of the modern world, especially as cell phones are increasingly used for internet connectivity, and more and more devices and appliances come with Wi-Fi. As these technologies grow, so too does the demand for more wireless bandwidth. Not only are more devices being used at the same time, but there is an ever-increasing demand for faster data transfer rates. More bandwidth almost inevitably means higher carrier frequencies. To date, most widely-employed

technologies reside in the low GHz range or lower. With millimetre wave technology in the 10s of GHz already being explored, e.g., for 5G communication, it is natural to see the THz as one of the key frontiers for future communication networks. From a bandwidth perspective alone, THz will be necessary if we want to achieve wireless transfer rates in the 100 Gbit s^{-1} range.

One important challenge with THz communication is absorption in the atmosphere (particularly in moist air), which tends to increase at higher frequencies. This may set some constraints on how the technology can be used in practice, but nevertheless there are promising applications over the entire THz spectrum:

- long range (1–10 km) communication at 100–150 GHz,
- medium range (0.1–1 km) communication below 350 GHz,
- indoor communication (10–100 m) below 500 GHz, plus two windows between 600 GHz and 1 THz, and
- near-field communications (< 10 m) above 1 THz.

The continued development of devices in these frequency ranges will therefore be important for the next generations of communications devices. Features such as compactness and efficiency are, of course, desirable for widespread applicability, but so are fast switching rates and high coherence. Sufficient power to overcome atmospheric absorption is also important – something which has traditionally been challenging in compact THz sources.

2.1.2 Quantum information

Quantum information has seen an enormous amount of development in the last few decades, with promises such as better data encryption and exponential speedups for certain computing problems. There are many challenges in this field, but from the perspective of photonics some of the hardest challenges involve creating "quantum" states of light, such as single photons or entangled photon pairs. Single photons, for example, are a crucial component of quantum key distribution, where unintentional generation of multiple photons can undermine the security of the system. Much progress has been made in this direction (e.g., [6–8]), but new sources of single-photons or entangled-photon sources are still of great interest at any frequency – THz included. As will be discussed later, THz polaritons may be particularly interesting from this perspective, due to the presence of vacuum state photons in the ultra-strong coupling regime [9].

2.1.3 THz imaging and spectroscopy

See, e.g., Refs. [1], [2] (Ch. 7), [4], and [10] for further discussions.

THz spectroscopy and imaging covers perhaps the broadest range of applications. Loosely speaking, the common thread in these applications is to probe a system with THz light and study how it reacts.

One reason this is such a fruitful research area is that many materials and systems have unique signatures in the THz frequencies. In medical research, for example, the THz can be useful for studying behaviours of biological molecules and drugs, such as their intermolecular vibrations. In atmospheric science and astronomy, THz systems can be tuned to detect certain gas species. In security applications, the THz can penetrate through packaging (while being safer than X-ray) to detect illicit materials such as

narcotics and explosives, which have unique signatures in the THz. These applications are already being developed and deployed, typically employing continuous-wave THz sources to probe the frequencies specific to a given material.

THz spectroscopy is certainly not limited to single-frequencies, though, as THz time-domain spectroscopy (TDS) allows for the study of ultra-fast processes via picosecond-scale pulses. THz TDS becomes a powerful tool for condensed matter research, where many scattering processes take place on picosecond timescales, and can be excited by electromagnetic waves. Even more sophisticated methods can be created by combining THz TDS with other techniques; for example, its combination with scanning tunnelling microscopy (STM) allows for "movies" with nanometre spatial resolution and picosecond temporal resolution [11]. There are applications outside the research lab as well – for example, local Waterloo company TeTechS is employing THz TDS to measure the wall thickness of plastic bottles on production lines.

2.2 Challenges in THz devices

Although many of the applications above have started to be implemented as THz technology development has accelerated in the last few decades, many challenges and limitations still remain. Here, we will focus on THz sources (rather than detectors) in particular. Even with sources, the challenges are certainly not universal, as each application comes with its own unique demands on device performance. Nevertheless, some general features which are often desirable in THz sources include

- high power output,
- high efficiency,
- coherence,
- tunability and/or good coverage of the THz range,
- room temperature operation,
- compactness, and
- low cost.

For broadband time-domain sources, we would typically want a large bandwidth or, equivalently, a sharp temporal resolution. For continuous wave sources, we would typically want a sharp frequency domain profile, and tunability may be important.

Many of these features have been challenging to achieve in the THz – much more so than in other, more well-developed spectral regions. It remains especially difficult to create versatile sources which cover many of these features at once. For example, one might have a high-power source which is too bulky and expensive for bench top applications in a laboratory. Nevertheless, many types of THz sources are actively under study, and are improving every year.

In the remainder of this chapter, we will survey some key approaches to generating THz light, discuss at a high level how they work, and provide some discussion of their current strengths and weaknesses. Again, this will not be an exhaustive survey, and the strengths and weaknesses of devices are always changing rapidly as new advances are made. Nevertheless, this will give us a small appreciation for the challenges in THz device design and will provide some context for our later discussion of THz intersubband polaritons.

2.3 Broadband THz sources

With broadband THz sources, the goal is typically to produce a short, high-power burst of THz radiation. By definition, this pulse is short-lived in the time-domain (on the scale of picoseconds), and therefore has a large bandwidth in the frequency domain. Currently, two prominent classes of broadband sources are photoconductive antennas and optical rectifiers.

2.3.1 Photoconductive antennas

Very roughly, a photoconductive antenna converts a narrowband optical pulse into a broadband THz pulse. An ultra-short burst of high-frequency light is shined onto an active material attached to an antenna. This light generates electron-hole pairs, which are then swept away by an applied bias voltage, creating a current in the antenna. This current, in turn, radiates THz light. (For a more thorough description see, e.g., Ch. 3.2 of [2]. This chapter is the source for the results below.)

To operate a photoconductive antenna, we first need a source of ultra-short high-frequency pulses. Fortunately, this is relatively easy to accomplish with, e.g., a Ti:sapphire laser, which can produce pulses of 800 nm light as short as 10–100 fs. These pulses might typically be of a few milliwatts.

Unfortunately, photoconductive antennas are not very efficient at converting this pulse into THz light. Typically, the THz output is on the order of a few microwatts, corresponding to a conversion efficiency on the order of 0.1 %. The overall device efficiency will be even worse than this, after accounting for the efficiency of the input laser, and losses due to dark current. The power level of a few microwatts is not huge, and it is primarily controlled by the applied bias voltage. More voltage means more output power, but too much voltage will destroy the device. The existing power levels can already be sufficient for spectroscopy applications, but there is always a push for more power to give better signal-to-noise ratios.

Another important property of photoconductive antennas is their bandwidth. (For time-domain spectroscopy, higher bandwidth means better temporal resolution.) Photoconductive antennas are typically able to achieve pulses with bandwidth of a few THz, covering most of the THz range. Photoconductive antennas also operate at room temperature, and they can be fabricated by standard techniques for semiconductor devices. The photoconductive antennas themselves are quite compact, but they require an external source to supply ultra-fast optical pulses, and this source may be bulky.

2.3.2 Optical rectifiers

Optical rectifiers are similar to photoconductive antennas in that an ultra-short optical pulse is converted into a THz pulse. However, the underlying mechanism is slightly different. Optical rectifiers rely on non-linear materials with multi-photon scattering processes. Higher-order effects allow one high-energy input photon to be converted into multiple low-energy output photons.

The key difference compared to a photoconductive antenna is that optical rectifiers do not require a high bias voltage. This is good, since breakdown due to the bias voltage is what limits the power output of photoconductive antennas. This allows optical rectifiers to obtain much higher power levels – limited instead by the available power of the optical input source, and the amount of input power that the nonlinear material can handle before breakdown. On the other hand, optical rectification relies on second-order scattering processes, so the conversion efficiency will typically be quite poor – even worse than photoconductive antennas. A typical value conversion efficiency value is ∼0.01 % [2]. (And again, this does not include losses in the input source and other optical components.) As with photoconductive

antennas, optical rectifiers require a combination of sources and components, meaning that they do not tend to be especially compact.

2.4 Continuous wave sources

An ideal continuous wave source would produce a pure sine wave in the electromagnetic field. In practice, this means a signal which is sharply-peaked in the frequency domain and long-lived in the time domain (long enough to include many oscillation periods which all have the same amplitude and frequency). This sort of source is useful for certain imaging and spectroscopy applications, as well as for fundamental science experiments where a well-defined frequency/energy is required. There are numerous continuous-wave THz sources, some of the most popular of which are discussed below.

2.4.1 Photomixers

A photomixer is essentially the continuous wave version of a photoconductive antenna (Section 2.3.1). Again, we have an active region attached to an antenna with a bias voltage applied. However, in this case, we shine two continuous beams of light at slightly different frequencies ω_1, ω_2, such that the difference frequency $\omega_2 - \omega_1$ lies in the THz range. Essentially, these two input beams will interfere with each other, resulting in a THz beat frequency. This generates a current in the photoconductive antenna which oscillates at the beat frequency, generating a continuous THz wave.

The properties of these devices will be similar to the photoconductive antennas used to generate broadband pulses. Low power output is a key disadvantage of this method, with good values being on the order 1–10 µW in the upper end of the THz range ∼1–3 THz [1]. Conversion efficiency is even worse in continuous wave mode versus broadband pulse mode, on the order of 10^{-5} or 10^{-6}. Both the power and efficiency drop off sharply as the emission frequency increases. One advantage of these devices, though, is that they can be easily tuned by changing the input optical frequencies ω_1, ω_2 (and tunable optical lasers are readily available). There is just a small trade-off here, in that the broadband antennas required for wide tunability will tend to give even weaker power output.

2.4.2 Difference frequency generation

Just as a photomixer is the continuous wave version of a photoconductive antenna, difference frequency generation is essentially the continuous wave version of optical rectification. Two input beams at optical frequencies ω_1, ω_2 are shined on a non-linear crystal to produce a THz beam of frequency $\omega_2 - \omega_1$.

As with optical rectification, this process is able to supply large power outputs at room temperature, on the order of 1 W. This is better than almost any other THz source. However, the conversion efficiencies are quite poor (again, because it relies on second-order photon scattering processes), and the overall device efficiency will be even worse. Further, these devices are table-top size, and not especially compact. These devices are widely tunable over the entire THz range, requiring just a tunable optical source. On the other hand, the output coherence is not particularly good and the bandwidth is not especially narrow.

2.4.3 Gas lasers

Gas lasers were one of the earliest type of lasers. A gas is confined in an optical cavity and pumped either electrically or optically. It can then lase, at a frequency which depends on the atomic transitions of the

particular gas being used. Certain gases will lase in the THz range, yielding power outputs of ∼10–100 mW [2]. There are a number of downsides to these devices though. For one thing, they are poorly-tunable, and it is not possible to design the laser to operate at any desired frequency – only specific frequencies can be obtained, depending on the gases available. Also, the efficiency does not tend to be good (on the order of 0.1 % [12]), and cooling is required to deal with the excess heat generated. Gas lasers further tend to be bench-scale systems which are not particularly compact.

2.4.4 P-type germanium lasers

P-type germanium lasers are able to achieve broadly-tunable THz output with high power outputs greater than 1 W. The catch is that their operation relies on Landau levels, which means that they require strong applied magnetic fields ∼1 T, as well as cryogenic cooling. While the devices themselves can be quite small, these requirements make it difficult to produce a complete package that is compact and efficient.

2.4.5 Microwave multiplication

Borrowing from microwave techniques, it's possible to multiply microwave frequencies using a Gunn diode or tunnel diode. This can be used to achieve compact, room temperature THz output with a narrow bandwidth ($\Delta\omega/\omega \sim 10^{-6}$) [2]. The power level drops off as the frequency increases, yielding ∼10–100 µW at around 1 THz.

2.4.6 Backwards wave oscillators

Backwards wave oscillators use electrons from a heated cathode to create electromagnetic waves whose group velocity is in the opposite direction to their phase velocity (backwards waves). The setup works better in the microwave regime, with power dropping off rapidly as the frequency increases. But, they can provide widely-tunable output up to ∼1 mW at ∼1 THz. Unfortunately, these are bulky and sophisticated systems. They require large magnetic fields ∼1 T, high cathode temperatures ∼1200 °C, and a high vacuum $\sim 1 \times 10^{-8}$ Torr [2].

2.4.7 Free electron lasers

Free electron lasers can produce high power light over a huge range of frequencies from microwave to x-ray. In the THz, they can produce tunable pulsed outputs with peak powers in the megawatt range. The catch is that a free electron laser is a large, expensive facility, requiring a team of people to operate it [2].

2.4.8 Quantum cascade lasers

Quantum cascade lasers (QCLs) are compact solid state lasers which emit in the infrared and THz range. Typical semiconductor lasers use transitions across the band gap to emit light, but even small band gap like InAs and InSb have band gaps of several hundred meV, well above the 1–10 meV of THz photons. So, rather than interband transitions, QCLs use intersubband transitions between the energy levels in a quantum well, allowing them to emit at lower frequencies like infrared (IR) and THz. QCLs were first demonstrated in 1994 [13] in the IR region. Since then, the IR QCLs have seen a great deal of success, with

high power outputs, good efficiencies, and room temperature operation. IR QCLs have even started to become commercially available.

THz QCLs, on the other hand, have been more challenging. Since their initial demonstration in 2002 [14], one of the biggest challenges has been temperature performance. Intuitively, it is not hard to see why this might be the case: to emit a THz photon, the spacing between the upper and lower lasing levels is necessarily on the order of 1–10 meV. On the other hand, at room temperature the thermal energy is $k_B T \approx 26\,\text{meV}$, which is enough to excite electrons to higher energy levels and act against population inversion. This does not put a fundamental limit on the operating temperatures, but it is nevertheless believed that thermal excitation at higher temperatures leads to parasitic tunnelling and non-radiative depopulation mechanisms [15, 16]. And, to make things worse, the lasing process itself tends to make the active region hotter than the surrounding temperature, meaning that the maximum operating temperature is reached even sooner.

To overcome this last problem, QCLs are often operated in "pulsed" mode, so that less time is spent lasing. This minimizes the heating of the active region. Note, however, that these "pulses" are still narrowband pulses, different from the broadband pulses used in time-domain spectroscopy. E.g., in [17] the pulses were 300 ns long, corresponding to 900 000 periods of a 3 THz wave. So the pulses will still be relatively sharp in the frequency domain.

Temperature performance

Using pulsed mode, the current record for highest operating temperature in a THz QCL stood at 200 K [17] for nearly a decade, only very recently reaching [18] 210 K and then 250 K [15]. These recent achievements in particular were a milestone, as they allowed the laser to be cooled thermoelectrically rather than cryogenically, making for much more compact and portable devices. This progress is certainly promising, but it remains to be seen how easily these new results can be replicated: aside from the active region design, the molecular beam epitaxy (MBE) growth and device fabrication quality may be important factors in the temperature performance. If the results can be replicated, then room temperature operation will be the next big milestone.

Still, it's worth noting that these record high operating temperatures come at a price. For one thing, high temperature operation tends to be achieved at "sweet-spot" frequencies around 3–4 THz, and the maximum operating temperatures are lower at higher or lower THz frequencies [16, 19]. For another thing, record-high temperature operation typically comes with severely reduced power output. While the device in Ref. [15] does lase at 250 K, it produces very little power. Only in the range 210–235 K can it produce the milliwatts of power required for spectroscopy applications. Nevertheless, these new results are very promising.

Power output

Currently, THz QCLs can achieve maximum power outputs of around 1 W for pulsed operation and 100 mW for continuous wave operation [19]. However, again, this sort of power is not achievable at high operating temperatures – only under cryogenic cooling. At the peak temperatures discussed above, the power outputs are more like ∼1 mW for pulsed operation [15] and ∼0.1 mW for continuous wave operation [20]. Furthermore, power tends to drop off significantly with lower-frequency QCLs [21].

Frequency range and tunability

THz QCLs cover essentially the top end of the THz range from ∼1–5 THz. This can be pushed lower with strong applied magnetic fields, but that has downsides from the perspective of compactness and efficiency. Also, as noted above, THz QCLs tend to perform better at the higher end of the THz range, near ∼4 THz. Power outputs, efficiencies, and maximum operating temperatures all get worse as the frequency moves closer to ∼1 THz. Tunability is achievable, though still a work in progress. Currently, THz QCLs can achieve continuous tunability over a range of a few hundred GHz [19].

Efficiency and threshold current

Efficiency is a bit complicated with QCLs. They tend to have relatively large threshold currents on the order of 0.1–1 kA/cm^2, so a lot of power is wasted before lasing even begins. However, once the lasing threshold is reached, the differential efficiency may be quite reasonable, with typical values being in the range of 100 mW/A. Still, the large threshold current means that wall plug efficiencies are at best around 1 %. And further, values like 1 % are only attainable at cryogenic temperatures. Since "wall plug efficiency" does not include the cost of cooling the device, the real efficiency may be much lower than this. Also, unsurprisingly from the above discussion, efficiencies will tend to be worse at higher temperatures and lower frequencies.

Beam shape

One challenge with THz QCLs is divergence of the beam. Even in single-mode operation, the beam may diverge in the far-field due to the sub-wavelength nature of the device. However, it is possible to reduce divergence to ∼10° × 10° [22].

Summary

As seen above, the performance of state-of-the-art QCLs varies quite a bit depending on which characteristic is being pursued. For example, while high power is achievable, it is only achievable with cryogenic operating temperature, and at the high end of the THz range. Similar trade-offs apply for the other characteristics. Still QCLs are arguably one of the most promising sources for the higher-end of the THz regime. They are under active development, and new milestones are achieved each year.

Operating temperature has long been one of the most serious challenges for these devices. The very recent progress has been quite promising, putting thermoelectrically-cooled THz QCLs within reach. Still, it may be some time before the "holy-grail" is reached: room temperature THz QCLs with enough output power for applications.

2.5 Final observations

Certainly, much more could be said about these THz sources and the growing field of THz technology in general. Many challenges will need to be tackled in the coming decades [4]. The existing THz sources have already begun to open up many applications, but there is still significant room for improvement. THz QCLs are one of the most promising THz emitters, but they are still not as mature as their IR counterparts. Room temperature operation remains impossible to date, and wall plug efficiencies remain low. Much work remains to create affordable, compact THz sources with significant output power.

This sets the stage for the remainder of this thesis, in which we develop techniques related to intersubband polaritons at THz frequencies. While the study of practical THz polaritonic devices is still in its infancy, these devices exploit the physics of strong light-matter coupling – something which has not traditionally been applied to optoelectronics. Developing a better understanding of the underlying science, with an eye towards practical devices, could form a useful part of the road towards mature THz technology.

Chapter 3

Background: exciton polariton lasers

Polariton lasers are a fascinating application of condensed matter physics to the development of practical devices. Their basic idea is to provide the same thing as a conventional laser – coherent light – but with distinctly different underlying physics. Conventional lasers, although they come in many different forms, typically rely on two of the same key ingredients: *population inversion* and *stimulated emission*. In the active region of a conventional laser, energy is pumped in to excite electrons to higher energy levels – when enough electrons are excited, this is called population inversion. When photons of an appropriate frequency are introduced, the electrons undergo stimulated emission, creating more and more photons. This amplification of light by stimulated emission, usually performed in a resonant electromagnetic cavity, is the fundamental process that underpins conventional laser action.

Polariton lasers, on the other hand, do not use population inversion or stimulated emission. They instead rely on a quasiparticle process analogous to Bose-Einstein condensation. Using a different underlying approach to generate coherent light – it is believed – may lead to practical advantages such as ultra-low lasing thresholds. In this chapter we will give an overview of the idea of polaritons in general, and look at some of the development of polariton lasers in the last few decades. Although the later chapters will focus on intersubband (ISB) polaritons, this chapter will primarily focus on exciton polaritons, which have been much more thoroughly studied. They form the foundation on which more recent studies of ISB polaritons have been built.

3.1 What are polaritons?

Polaritons are quasiparticles which arise from the strong interaction between light and matter. Loosely, the idea is that when light and matter interact strongly enough, the natural excitations of the system become mixture states: part light and part matter. This idea can be made more mathematically precise, and there are several ways to do so, both semiclassical and quantum mechanical [23]. Both approaches are important for calculations and modelling, but the quantum mechanical picture is, in some ways, more intuitive.

In the quantum picture, polaritons arise when the light-matter interaction is too strong to be treated perturbatively. In most of the situations we encounter in physics and engineering, light-matter interaction is weak, and we can treat it perturbatively using Fermi's golden rule. Fermi's golden rule allows us to calculate, say, the probability of an atom in free space absorbing a photon in a given timeframe. However,

we could create situations where the light-matter interaction is much stronger – say, by putting the atom in a resonant cavity. If the interaction is strong enough, we can no longer apply Fermi's golden rule, and we must turn to a more direct method to diagonalize the Hamiltonian. As we will see, the eigenstates of a such a system turn out to be superpositions of light and matter excitations. These superposition states are what we call polaritons.

3.1.1 Quasiparticles

More generally, polaritons are a type of *quasiparticle*, which is a widely useful concept in modern condensed matter physics. One of the overall goals of condensed matter physics is to describe large systems of many interacting quantum particles. Needless to say, this can be an extremely difficult task. The wavefunction of N spinless interacting particles lives in $3N$ dimensional space, so the problem grows exponentially in difficulty as more particles are added. The basic idea of quasiparticles is to simplify this problem by finding a more convenient way of looking at a system. Often, it turns out that a complicated system of interacting particles can be re-written as a simpler system of non-interacting quasiparticles. Or, at least, a system of strongly interacting particles can be re-written as a mathematically equivalent system of weakly interacting quasiparticles.

The idea of quasiparticles should be familiar to anyone who has studied semiconductor physics, although perhaps not under the same name – electrons and holes in semiconductors are, in fact, prime examples of quasiparticles. We know that in reality, a semiconductor is made up of interacting electrons, protons, and neutrons. In principle, we could model semiconductors directly in terms of these particles, but this is an incredibly arduous task in practise. Instead, by re-writing the problem in terms of fictitious "quasi"-particles – holes and conduction electrons – we obtain a much more manageable picture. This core idea is much more broadly applicable, and there are many other types of quasiparticles in solid state physics such as phonons, excitons, magnons, plasmons, and – the subject of interest here – polaritons. All of these are just simpler ways to describe the complicated interactions of the full many-body system. The key physics of the many-body system is folded into the properties of these particles, such as their effective mass (or, more generally, dispersion relations) and their interactions with other quasiparticles.

To make this more concrete, we can consider a very simple model of the coupling between two types of particles, a and b. For the purposes of this discussion, we can assume that a represents photons and b represents some matter excitation like an exciton or a plasmon. We can imagine a simple interaction between these types of particles described by the following second-quantized Hamiltonian:

$$H = \hbar\omega_a a^\dagger a + \hbar\omega_b b^\dagger b + \frac{\hbar g}{2}\left(a^\dagger b + a b^\dagger\right) \tag{3.1}$$

Here $\hbar\omega_a, \hbar\omega_b$ are the energies of a single photon and matter excitation, respectively, and $\hbar g$ is the coupling energy. $a^\dagger$ and $b^\dagger$ are the creation operators for photons and matter excitations, respectively. So, if $|0\rangle$ is the vacuum state with no particles, then $a^\dagger|0\rangle$ will be the state with exactly one photon, and $b^\dagger|0\rangle$ will be the state with exactly one exciton (up to normalization). For simplicity, we are assuming here that these particles can only exist in one possible state, however, the same approach could be extended to include energy dispersions by summing over the wave vectors, $\mathbf{k}$.

As a starting point, if we take away the coupling term ($g = 0$), then this Hamiltonian becomes quite simple: $H = \hbar\omega_a a^\dagger a + \hbar\omega_b b^\dagger b$. Since $a^\dagger a$ and $b^\dagger b$ are the number operators of photons and matter excitations respectively, the eigenstates of the Hamiltonian are just states with exactly n_a photons and n_b matter excitations. These states have energy given by $E_{n_a,n_b} = n_a\hbar\omega_a + n_b\hbar\omega_b$. Further, since these states

are eigenstates of the Hamiltonian, they will be stationary states. Thus, if we start with a specified number of matter excitations and photons, that will not change. When no interaction is present, matter excitations won't be converted into photons or vice-versa.

However, if we turn on the interaction term by setting $g \neq 0$, things get more interesting. States with exact numbers of matter excitations and photons will no longer be stationary states. Matter excitations can now be converted into photons and vice-versa, as we might expect when light-matter interaction is present. It turns out that in this case the stationary states (eigenstates of the Hamiltonian) will be *superpositions* of different numbers of matter excitations and photons. This can be shown by defining a new set of creation operators, which create superpositions of matter excitations and photons:

$$p_{\pm}^{\dagger} = \alpha_{\pm} a^{\dagger} + \beta_{\pm} b^{\dagger} \tag{3.2}$$

It is possible to choose the coefficients $\alpha_{\pm}, \beta_{\pm}$, in just the right way to collapse the interaction Hamiltonian (3.1) into a diagonal form:

$$H = \hbar\omega_{+} p_{+}^{\dagger} p_{+} + \hbar\omega_{-} p_{-}^{\dagger} p_{-} \tag{3.3}$$

(By combining this expression with Eqs. (3.2) and (3.1), it is not too difficult to work out exactly what the coefficients $\alpha_{\pm}, \beta_{\pm}$ and the quasiparticle energies $\hbar\omega_{\pm}$ must be.) So, by defining new creation operators which are superpositions of the original creation operators, we have transformed an interacting Hamiltonian in to a *non*-interacting Hamiltonian. The picture of interacting photons and matter excitations has been transformed into a picture of non-interacting light-matter particles, which we call polaritons. So we must pay the cost of introducing new types of particles, but there is a clear gain in terms of simplifying the problem.

In some sense, then, the existence of polaritons can be seen as just a mathematical "trick" – we are simply switching to a basis in which the Hamiltonian matrix is diagonal. In principle, everything could be equally well described with the original exciton/photon picture without referring to polaritons at all – these are just two different ways of looking at the same problem. However, quasiparticle pictures like this one can simplify a problem considerably, making them powerful tools for understanding physical phenomena.

3.1.2 Exciton polaritons

The above example illustrates the core idea of polaritons – that they are a superposition of a material excitation and a photon – but there is obviously much more that can be said. To actually create polaritons in practice, the most popular method has been to use quantum well (QW) excitons in a microcavity. An example of such a structure is shown in Fig. 3.1. If designed appropriately, this structure will create cavity photons and QW excitons which are strongly coupled together, as first demonstrated in 1992 [24].

It is useful to look at the resulting dispersion diagram, an example of which is shown in Fig. 3.2. In the absence of coupling, the exciton and photon modes can overlap with each other. But, when coupling is added, there is an anti-crossing behaviour called Rabi splitting. When the photon and exciton energies are close together and the coupling strength is high, the polariton energies are pushed apart from each other. In this case, there is also a strong mixing of the modes. The polariton becomes a near-equal superposition of exciton and photon. So, for example, while the exciton has a very high effective mass (it's dispersion is almost flat), the lower polariton has a small region of low effective mass inherited from the cavity photon.

Though not captured in the figure, loss is an important mechanism which must be accounted for whenever we are discussing polaritons. In real systems, cavity photons and QW excitons will both have

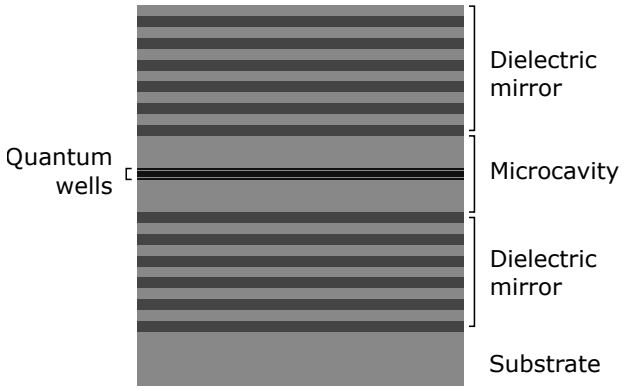

Figure 3.1: An example structure for creating exciton-polaritons. This is a planar structure which could be grown, for example, with molecular beam epitaxy in III/V semiconductors. Dielectric mirrors – alternating $\lambda/4$ sections of materials with different refractive indices – form the cavity and confine the light. By varying the cavity thickness, the resonant frequency of the photon modes can be tuned to the exciton energy. Quantum wells at the centre place excitons at the location of the highest electric field strength.

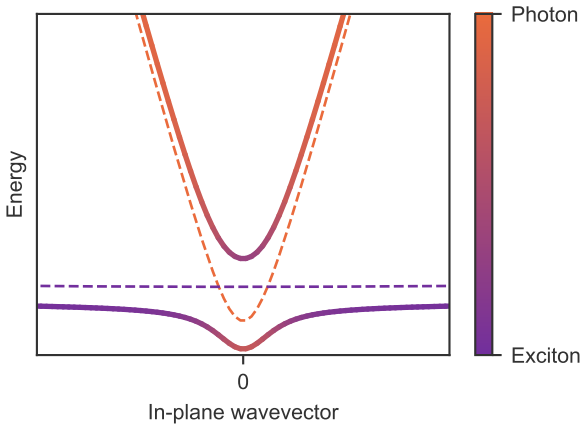

Figure 3.2: A qualitative picture of polariton dispersion and mixing (scales are in arbitrary units). On the x axis we have the in-plane wave vector, $k_{||}$, for the layered structure. The dashed lines show the bare (non-interacting) cavity photon and exciton dispersions. Note that the exciton dispersion appears flat because its effective mass is many orders of magnitude smaller than the photon effective mass [25], and we are only looking at a small section of the Brillouin zone near $k_{||} = 0$. The solid lines show the upper- and lower-polariton dispersion. The colour indicates the mixing fraction – how photon-like or exciton-like the polariton is. Near $k_{||} = 0$ there is strong mixing, while at higher k values lower-polaritons are essentially just excitons and the upper-polaritons are essentially just photons. Maximal mixing (50:50) occurs at the crossover between the exciton and photon dispersions.

finite energy broadening parameters, γ_a and γ_b respectively. These arise from, e.g., the lifetime of the cavity photon before it escapes the cavity, and the decoherence of excitons due to phonon scattering. These losses are, in fact, what separates strong coupling from weak coupling. If $\hbar g$ is the light-matter coupling energy, then the condition for strong coupling becomes:

$$g > \gamma_a + \gamma_b \tag{3.4}$$

That is, the coupling rate must exceed the decoherence rates of the underlying particles. This can be derived from the simple model above by including a small imaginary part in the energies of the original particles – or, see Ref. [25] for a simple derivation using 2×2 matrices. To achieve strong coupling, then, it is important to have high-quality cavities and quantum wells to minimize the decoherence rates γ_a, γ_b. Additionally, overlap between the electric field and the exciton dipole moment should be maximized to increase g.

So, to summarize, polaritons appear when there is strong coupling between light and matter – strong enough to overcome the losses present in the system. In that case, the stationary states of the system become superpositions of light and matter. By treating these superpositions as new particles called polaritons, we gain a very convenient picture of how light and matter behave in the strong coupling regime.

There is much more that can be said about exciton polaritons – some of which will be discussed below – but this gives a very brief introduction. Further introduction can be found in a variety of textbooks [23, 26, 27] and review papers [25, 28–32].

3.2 Exciton polariton condensation

Polaritons have some interesting properties that have led to their intense study over the last few decades. One thing that particularly caught interest was their utility as a platform for studying phenomena related to Bose-Einstein condensation. Polaritons are bosons (strictly, composite bosons [33]), as they are made up of photons and material excitations which are each bosons themselves. It is well known in statistical physics that systems of bosons can form Bose-Einstein condensates at low enough temperatures – a phenomenon in which a significant fraction of the particles congregate into a single quantum state. Traditionally, Bose-Einstein condensation is a result of equilibrium thermodynamics, but it turns out that similar processes can happen out of thermodynamic equilibrium as well. In the context of polaritons, we refer to these processes as polariton condensation [30].

One thing that makes polaritons so attractive as a platform for studying these phenomena is their hybrid light-matter nature. Their photon fraction gives them a very small effective mass, which allows them to condense at relatively high temperatures. Polariton condensation has been observed even at room temperature [34], while atomic systems can only Bose condense at extremely low cryogenic temperatures (e.g., early demonstrations [35, 36]). The photon fraction also provides convenient way to probe and observe polariton condensates – when a polariton decays and emits a photon from the surface of the sample, the angle and energy of that photon indicate the energy and k of the polariton that emitted it.

The exciton fraction is equally important, as it provides a way for polaritons to interact and "thermalize" to form a condensate. The small effective mass of photons is useless for condensation if there are no scattering mechanisms which allow them to reach the ground state within their lifetime. Photons do not interact with each other, so the only way to get them to thermalize is through interaction with a material. Creating polaritons is not the only way to do this – for example, it has been shown that photons

can thermalize via weak interactions with dye molecules [37] – but the strong interaction with excitons remains a fruitful mechanism for study. The exciton fraction of polaritons supplies them with numerous scattering mechanisms – polariton-phonon, polariton-polariton, polariton-electron, and others – which can work together to bring polaritons towards the ground state and build up the presence of a condensate.

A key idea here is the notion of *stimulated scattering*. Because polaritons are effectively bosons, their scattering rate towards the ground state is multiplied by $(1 + \langle N_0 \rangle)$ where $\langle N_0 \rangle$ is the number of polaritons occupying the ground state. If there are no polaritons in the ground state ($\langle N_0 \rangle = 0$), then any scattering towards the ground state will happen at its spontaneous scattering rate. However, if there is already a population of polaritons in the ground state, the scattering rate is enhanced. This can lead to a positive feedback loop which allows a large buildup of polaritons in the ground state. It is this positive feedback loop of stimulated scattering which is at the heart of polariton condensate formation.

There has been a significant amount of experimental work studying these condensates. An exciton polariton condensate was first clearly demonstrated in 2006 [38], although there were earlier indications of stimulated scattering [39]. In the last two decades, there have been many fascinating experiments probing aspects such as superfluidity and vortices – see e.g., [30] and the references therein.

3.3 Exciton polariton lasers

Polariton condensates are already worth studying just from the perspective of fundamental physics. But it was understood early on that they could also have practical applications through the so-called "polariton lasing" mechanism [40]. Polaritons, because of their cavity photon fraction, can spontaneously decay to emit a photon into free space. Importantly, if the polaritons are in a condensate – macroscopically occupying the same quantum state – the light they emit should ideally inherit coherence from the condensate, producing laser-like light.

This is particularly interesting because no population inversion is required to create a polariton condensate. Population inversion is a key requirement for conventional lasers. Enough electrons must be excited into higher energy levels to create a gain medium via stimulated emission. Creating this population inversion requires some threshold input power, below which lasing will not occur. Polariton lasers, on the other hand, only require that the scattering rate towards the ground state exceeds the polariton lifetime. This can be a much less stringent requirement than population inversion, in theory allowing polariton lasers to have much lower thresholds than their conventional counterparts.

Development of these devices is still underway, but important steps have been made in the last 20 years. Optically-pumped devices were demonstrated in the mid-2000s, first at cryogenic temperatures [38] and later at room temperature [34]. Electrical injection proved more difficult, and was finally demonstrated in microcavities in 2013 by two separate groups [41, 42]. A report of room-temperature, electrically-injected polariton lasing followed the next year [43].

Still, challenges remain. Though these devices do provide relatively low thresholds, they are not yet living up to their theoretical potential, and lower thresholds can be found in photonic crystal quantum dot lasers [44]. This is attributed to the so-called "bottleneck" effect, where the acoustic phonon scattering rate decreases near the inflection point of the lower-polariton dispersion. At the same time, the photon fraction increases and the lifetime decreases, so more power is required to overcome the radiative loss of non-condensed polaritons. Improved cavity qualities and/or optimized scattering mechanisms may be needed to push thresholds lower.

Another recent focus has been the coherence of the output light, which would be important for a practical laser device. Many demonstrations of polariton lasing to date have exhibited intensity fluctuations

and somewhat lacklustre phase coherence compared to conventional lasers. Recent effort has gone towards isolating the reasons for this, and robust second-order coherence has now been shown in an optically-pumped low-temperature polariton laser [45].

Polariton devices are still in their infancy, but the future looks promising. And, as far as polariton devices in general go, lasers are just the tip of the iceberg. Many other applications are being explored such as optical logic gates or polariton quantum simulators. It remains to be seen which of these ideas will mature, but nevertheless exciton polaritons provide fascinating possibilities to apply condensed matter theory in the real world.

3.4 Exciton polaritons in the THz

In the remainder of this thesis, we will focus on ISB polaritons, which are particularly well-suited for the terahertz (THz) and infrared (IR) regions. As we saw in Ch. 2, the THz is an especially interesting frequency range in which to develop new devices. However, before we turn to ISB polaritons, it is worth noting that there have been some suggestions for THz lasers based on exciton polaritons. In some sense, exciton polaritons are a natural choice, because they have underpinned the bulk of the work so far on polariton condensates and optical polariton lasers. The theoretical and experimental methods are becoming ever more mature, backed by a significant body of literature.

Unfortunately, producing THz light with exciton polaritons is not quite as simple as optical light. They essentially suffer from one of the common difficulties of building semiconductor THz lasers – semiconductor bandgaps are just too large. The exciton polariton energy is essentially determined by the band-gap of the material employed (binding energy aside), which means that their energy will typically be on the order 1 eV. This is well above the THz regime of ∼0.1–10 meV. So, even in a narrow bandgap semiconductor like InAs (bandgap ∼354 meV), exciton polaritons will not emit THz photons upon decay. Because of this, the standard exciton polariton lasing mechanism cannot be used in the THz regime.

To overcome this, proposals for THz emission have often focused on inter-polariton transitions. For example, a THz photon can be released when an upper-polariton is converted to a lower-polariton [46], or when a $2p$-exciton is converted into a lower-polariton [47]. A slightly more sophisticated approach uses *dipolaritons* [48–50], which are a three-way mixture between a cavity photon, a direct exciton, and an indirect exciton. Again, though, the idea is to exploit the fact that the energy difference between two different polariton states can be in the THz regime, even though the polaritons themselves have energies in the optical or IR regime.

As with optical polariton lasers, the bosonic nature of polaritons provides an advantage here. In particular, if the final state is heavily occupied (i.e., a condensate exists), the scattering rate and the emission rate of THz photons will be enhanced. In the presence of a THz cavity, this could lead to a sort of double-stimulated emission, with the emission rate enhanced both by the number of THz photons in the cavity and by the number of polaritons in the final state. Alternatively, by relying on bosonic stimulation alone, it's possible that coherent THz light could be produced without even needing a THz cavity [47].

Of course, there are challenges with these approaches. One challenge is the selection rules, which prevent, for example, the optical transition from upper- to lower-polariton in symmetric quantum wells. In the approaches above, these selection rules are carefully avoided with asymmetric quantum well structures.

A more serious challenge is efficiency. In optically pumped systems, even if every exciton polariton produces one THz photon, the device efficiency is immediately limited to ∼1 % at best, as a ∼1 eV exciton polariton must be created for each THz photon (∼10 meV). Some attempts have been made to push the

conversion efficiency above unity, for example by using a cascade of transitions between equally-spaced polariton states in a parabolic quantum well [51–54]. This could improve the conversion efficiency by a factor of 10 or so (the number of levels in a parabolic well is limited by the materials used). However, in practice the conversion efficiency is still limited to well below unity as THz emission competes with optical emission and non-radiative scattering processes. Even under optimal configurations where non-radiative scattering is negligible, it may still require more than 1000 exciton polaritons to produce a single THz photon [55]. These factors alone are likely to push the device efficiency well below $\sim 10^{-5}$.

Finally, it is worth noting that these devices are still in the very early stages of their development. The papers cited above, as well as the efficiency values quoted, are all theoretical. It remains to be seen what challenges will appear when implementing these devices in practice.

Chapter 4

THz devices with intersubband polaritons

In Chapter 3, we introduced the notion of polaritons and polariton condensates. Our discussion drew mostly from research on *exciton* polaritons, which have received the most intense study. However, polaritons need not be based on excitons – they can in principle be formed from any type of material excitation that couples to photons. What we will focus on in the remainder of this thesis is the strong interaction between light and intersubband plasmons in a quantum well (QW): so-called *intersubband* polaritons (ISB polaritons).

In Ch. 5 we will look more closely at the modelling of intersubband (ISB) plasmons, the matter part of this polariton recipe. Essentially, though, an ISB plasmon is the quasiparticle associated with electron transitions in a QW. When we promote electrons from lower to higher subbands in a QW, this is equivalent to the creation of ISB plasmons in the quasiparticle picture. So ISB polaritons arise when light interacts strongly with the ISB transitions of a QW – for example, when we place a heavily doped QW into a resonant cavity.

ISB polaritons have attracted attention for both theoretical and practical reasons. One of their most obvious differences compared to exciton polaritons is that they occupy a lower end of the frequency spectrum – typically in the terahertz (THz) and mid-infrared (IR), as opposed to the near IR and visible frequencies of most exciton polaritons. The difference is more than just the operating frequency, however. One interesting feature of ISB polaritons which motivated much of their early study is the so-called ultrastrong coupling regime – distinct from the strong coupling regime seen with exciton polaritons. In exciton polaritons, the Rabi splitting (which is directly related to the light-matter coupling energy) is typically small compared to the original exciton and photon energies. In contrast, with ISB polaritons, the coupling strength can be an appreciable fraction of the original plasmon and photon energies. This is helped by the fact that the ISB transition and photon energies are relatively small.

The ultra-strong coupling regime is not just quantitatively distinct from the strong coupling regime, but there are qualitative differences in polariton behaviour as well. Perhaps most interesting is the possibility of virtual photons in the ground state of ultra-strong coupled systems, which could make them a potential platform for entangled photon sources [9]. This idea motivated much of the early experimental work, with Rabi splittings eventually reaching near-equality to the matter excitation energy [56]. This was accompanied by theoretical work, for example showing that the virtual photons are maintained in the presence of loss [57].

Other differences in ISB polaritons compared to their exciton-based counterparts make them interesting from a device perspective. One of their more subtle differences is the robustness of their bosonic

nature. Strictly speaking, exciton polaritons and intersubband polaritons are only approximately bosonic – being ultimately derived from photons (bosons) and electrons (fermions). They are instead *composite* bosons, which can behave differently from fundamental bosons [33]. In the exciton polariton case, these deviations arise prominently when the density of polaritons exceeds a certain value, known as the Mott transition. Roughly, once the excitons are dense enough to overlap spatially, the exchange statistics of their fermionic components becomes non-negligible, and they lose their bosonic character. In the case of polariton lasers, this essentially limits the maximum output power as the lasing mechanism breaks down above a certain density, and the limit is set by the material.

Intersubband polaritons, too, experience a loss of bosonicity above a certain density. Importantly, though, this transition can be engineered by increasing the doping density and the number of quantum wells in the system [58]. In the context of polariton lasers, this could mean that ISB polariton lasers are less limited in their output powers while still maintaining the advantage of a low threshold. For the THz, it has been suggested that output powers on the order of 10 mW may be achievable at room temperature [59]. If this can actually be achieved, this would compare very favourably with existing THz technology.

Of course, designing an actual laser with intersubband polaritons is not a trivial task. This field is still very young, and while there have been promising demonstrations such as electroluminescence at room temperature [60], a working laser device has not yet been demonstrated. In the following sections, we will discuss the existing state of the art and challenges.

4.1 The active region

To make a THz ISB polariton laser, we need to create polaritons and make them undergo stimulated scattering towards the ground state. To make ISB polaritons, we need strong coupling between ISB plasmons and a photon mode. Typically, this is done by placing an active region inside a THz photonic cavity.

The active region is typically made of a stack of multiple quantum wells, which are fairly heavily doped (often in excess of 1×10^{11} cm^{-2} or even 1×10^{12} per well). To obtain strong coupling, the quantum wells should be designed with an ISB absorption frequency tuned close to the electromagnetic cavity resonance. The strength of the absorption is controlled by (a) overlap between the quantum wells and the cavity mode and (b) the plasma frequency of the QW resonance. Ideally, the plasma frequency is proportional to the square root of the 2D doping density, which makes doping an important parameter for engineering the coupling strength. As we discussed in the last chapter, the basic criteria for strong coupling is that the interaction strength needs to exceed the losses present in the system.

So, to make a good active region, we need to have a plasmon absorption resonance in the frequency range of interest, and we would usually want to have a strong absorption resonance with a small linewidth. As we will see later, this can be particularly challenging in the THz. This challenge will, in fact, be the primary focus of the later chapters as we try to improve upon the active region's design and growth.

Fortunately, though, we do not need to delve into the details quite yet to understand the overall design of ISB polariton structures. It turns out that many of the most important aspect of devices can be modelled semiclassically, replacing the QW active region with an effective permittivity tensor. Calculating this tensor theoretically is no small task, and will be the focus of Chapter 5. We can assume for the purposes of device design, though, that the active region simply produces an absorption resonance with a certain centre frequency, width, and absorption strength.

As a starting point, based on the method of Załużny and Nalewajko [61], we can replace the QW stack

with a dielectric tensor whose growth-direction component is

$$\frac{1}{\epsilon_{zz}(\omega)} = \frac{1}{\epsilon_s}\left(1 + \frac{\omega_P^2}{\omega^2 - \omega_0^2 + i\gamma\omega}\right) \tag{4.1}$$

where $\epsilon_s (\approx \epsilon_{xx} = \epsilon_{yy})$ is the static dielectric constant of the QW stack. ω_0 is the centre frequency of the absorption, ω_P is the effective plasma frequency (related to the absorption strength), and γ is the linewidth. Note that this produces an anisotropic absorption, in accordance with the well-known polarization selection rules for QW.

In the simple case of a QW with a single active ISB transition, we can calculate the quantities ω_0 and ω_P in a straightforward way. $\hbar\omega_0$ is simply the ISB energy spacing of the QW, and ω_P is given by

$$\omega_P^2 = \frac{e^2 N_{2\mathrm{D}}}{\epsilon_0 \epsilon_s m^* L} \tag{4.2}$$

where $N_{2\mathrm{D}}$ is the 2D doping density per QW, m^* is the effective mass, and L is the total width of a single period (QW plus barrier). However, this single-transition assumption does not always hold, and will break down for many of the active regions considered in this thesis. This happens whenever more than one subband is occupied due to heavy doping and/or high temperatures. (THz QWs are particularly prone to this due to their small ISB spacing.) In those cases, ω_0 and ω_P can shift significantly from their nominal values, or even become ill-defined.

However, even in these cases where the single-transition assumption breaks down, the Załużny-Nalewajko formula, often remains a useful descriptive model. As long as we design the QW to produce one prominent absorption resonance, the effective dielectric constant can usually be fairly well-approximated by Eq. 4.1 for *some* values of $\omega_0, \omega_P, \gamma$. It will be much more difficult to predict what those quantities actually are for a given active region, but the important thing is that we can still use the mathematical form of Eq. 4.1 to study a wide range of polariton devices. By treating Eq. 4.1 as an empirical model, we can get an idea of what characteristics the active region absorption needs to have in order to reach certain goals like polariton scattering and lasing. Then, in later chapters we will return to the problem of how to actually design and grow an active region to achieve the values of ω_0, ω_P, and γ that we want.

4.2 The photon cavity

We will now turn our attention to the THz photon cavity, in which the active region will be placed. There is, of course, a plethora of possible cavity designs. To narrow them down, we can create a "wishlist" of features that would be desirable for THz polariton lasing. To create strong coupling and polaritons, we want

1. a photon resonance in the THz regime,
2. a long photon lifetime (a high Q-factor),
3. an electric displacement field $\mathbf{D}$ with a strong component in the growth direction z, and
4. good spatial overlap between $D_z(\mathbf{r})$ and the quantum wells.

For the formation of a polariton condensate and polariton lasing, there are a few more desirable features:

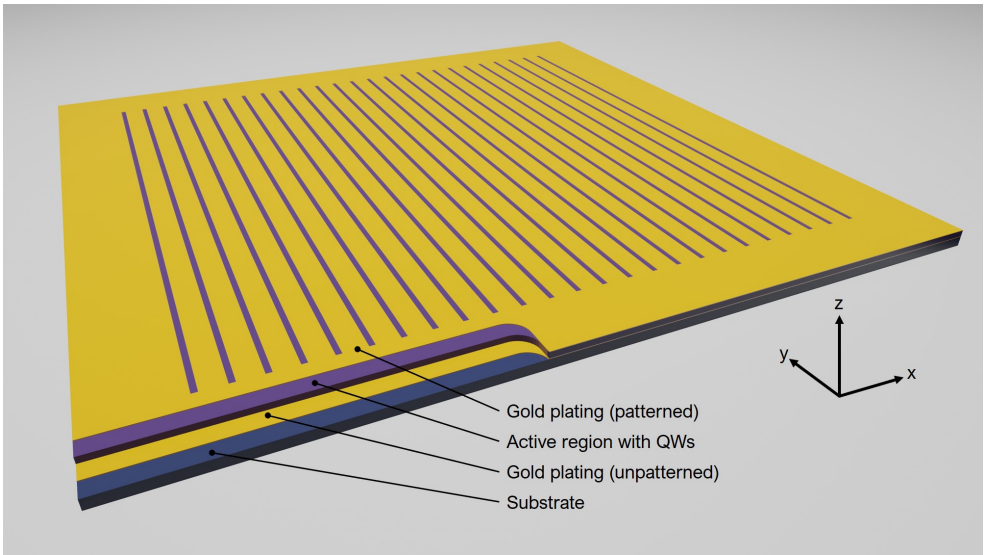

Figure 4.1: 3D example of a metal-insulator-metal (MIM) cavity section. (The cutaways on the bottom left would not be present in an actual device. They are just to show the different layers.) There are four layers, from top to bottom: the patterned metal (usually Ti/Au) with slits cut out, the active region, an unpatterned metal layer, and the substrate. The active region itself will typically comprise many layers, e.g. in the form of doped quantum wells. For THz frequencies, the thickness of the active region is typically up to approx. 10 µm, which is around the maximum thickness that can be feasibly grown with molecular beam epitaxy (MBE). The periodicity of the slits is tens of µm, meaning that the metal can be patterned using optical lithography. In reality, the total extent of the slits may be several millimetres in each direction – much larger than what is shown here.

6. a photon dispersion minimum with strong positive curvature (low effective mass),
7. the possibility of contacts for electrical injection, and
8. good coupling of cavity photons to the external photon modes.

Several cavity designs have already been used in intersubband polariton experiments. The simplest approach is to coat the top of the active region in a metal, and use refractive index contrast on the bottom to confine the light. This was used in early strong and ultra-strong coupling experiments [62–64]. A more sophisticated version patterns slits into the metal to allow illumination from the top rather than the side [65]. However, because dielectric confinement is used, the photon dispersion is not correct to satisfy requirement 6 above. So these cavities are not suitable for polariton condensate formation and lasing.

Another approach uses LC-resonators patterned onto the surface of the active region [66]. This cavity geometry is useful for strong and ultra-strong coupling [67, 68] and even for electrically-driven emission at room temperature [60, 69]. However, again, the (lack of) dispersion of these cavities makes them unsuitable for condensate formation.

One promising approach uses a MIM geometry, with the top metal layer patterned into rectangular slits or islands to form a photonic crystal [70, 71]. (See Fig. 4.1, for example.) These cavities have been used

for several strong coupling demonstrations [72, 73]. The TM_0 mode of these cavities can exhibit excellent electric field overlap, allowing for demonstrations of ultra-strong coupling [72, 73], including record-high coupling strengths at room temperature with the Rabi splitting reaching 73 % [74], and even 91 % [56] of the matter excitation energy. Additionally, by appropriate design of the photonic crystal, it is possible to make a dispersion minimum required for polariton condensation. Intersubband polaritons with a low effective mass have already been demonstrated in the mid-IR [75].

These cavities are also advantageous from the perspective of coupling to external cavity photons. One well-known challenge with intersubband transitions in a QW is that they will only couple to an electric field oriented in the growth direction [76]. This makes it difficult to excite intersubband transitions with normally-incident light, since the electric field is oriented in the wrong direction. These patterned MIM cavities, however, essentially "steer" an incoming beam, so that the TM modes can be excited by normally-incident light [71]. This makes it easier to perform, e.g., angle-resolved pump-probe measurements.

Further, it has been shown that etching the active region through the slits can provide even more control over the external coupling. A so-called "critical-coupling" condition can be met to allow perfect interferometric energy feeding into a sample [65]. At this same time, this process allows a tuning of the photon-to-plasmon mixing ratio so that the $\mathbf{k}_{\parallel} = 0$ polaritons have maximal mixing [77]. Dispersive MIM cavities are, therefore, a promising path for studying the phenomena relevant to THz intersubband polariton lasing, and potentially for lasing itself.

4.3 Strong coupling with MIM cavities

Let us now explore further the behaviour of strong coupling in MIM cavities. This will give us an idea of what sort of quality is desirable from the QWs in the active region. The discussion in [71] gives some intuitive properties of these cavities, but a full numerical model is important to calculate photon resonances, lifetimes, and coupling to the matter excitation.

As in the case of exciton-polaritons [23], there is more than one way to model the coupling between the cavity photons and the ISB plasmons in the active region. Some quantum electrodynamical models exist [78, 79], but these can be complicated to implement. Instead, we rely on a semiclassical approach, in which the active region QW stack is replaced by an effective anisotropic dielectric tensor. Here we assume the single-mode form of Eq. (4.1). Then, modelling the cavity simply requires an electromagnetic solver. For that, we will use the rigorous coupled-wave analysis (RCWA) method [80–83]. (The Matlab code used here was developed by Simone Zanotto and Jean-Michel Manceau.) This combination of the RCWA solver with an effective dielectric tensor has already proved successful in the mid-IR for predicting the reflectivity spectra of ISB polaritons in dispersive MIM cavities [75].

As an example system, we will look at a MIM cavity (similar to Fig. 4.1) designed to have a resonance around 2.8 THz with an upward-curved dispersion. The top and bottom metal regions are made of gold. The top grating has a periodicity of $\Lambda = 36\,\mu\text{m}$ and a duty cycle of 80 % (i.e., the metal strips are 28.8 µm wide). The active region is a total of 9 µm thick, assumed to comprise a 6 µm stack of AlGaAs QWs on the bottom and 3 µm of undoped GaAs. We assume that there is one QW every 150 nm, doped at $1 \times 10^{11}\ \text{cm}^{-2}$ per well, with an absorption centre at $\omega_0 = 3.0\,\text{THz}$. The QW permittivity is calculated according to Eq. (4.1), assuming a simple single transition.

With the RCWA code, we can simulate reflection off the MIM cavity at an incidence angle θ. We assume that the light is incident in the plane perpendicular to the slits (the x–z plane in Fig. 4.1), and that it is transverse magnetic (TM) polarized. (So the magnetic field is along the y axis, and the electric field lies in the x–z plane.) This sort of angle-resolved reflection experiment is useful because it can be directly

mapped onto a dispersion diagram for the quasiparticles in the cavity. The incoming photons have energy $\hbar\omega$ and an in-plane wave number of

$$k_{\parallel} = (\omega/c)\sin(\theta), \tag{4.3}$$

and these photons can only be absorbed by the cavity if they have an energy and wave number that corresponds to one of the modes of the cavity. Thus, by looking at the dips in the reflectivity as a function of ω and $k_{\parallel}$, we can see a map of the cavity dispersion. Since these reflectivity maps can be readily generated both experimentally and theoretically, they are a useful tool for understanding and designing polariton cavities.

To start, we look at the empty cavity dispersion in Fig. 4.2(a), where the QW doping has been removed. Note that there is forbidden region since external photons cannot exist with $|k_{\parallel}| > \omega/c$ regardless of the incidence angle θ. This defines the so-called "light cone," below which we do not plot any reflectivity. Above the light cone, though, we see two clear photonic modes, a lower-energy mode with a dispersion maximum around 2.5 THz at $k_{\parallel} = 0$, and a higher-energy mode a dispersion minimum around 2.8 THz at $k_{\parallel} = 0$. It is the latter mode that we are interested in for polariton scattering studies and condensate formation: a sample electric field pattern for this mode is shown in Fig. 4.3 for $\theta = 30°$. The Q factor of this mode (the ratio of its linewidth to its frequency) varies from around 25–30. This is quite low compared to the high-quality Bragg mirror cavities used in exciton-polariton experiments, however, due to the lower frequencies involved here, it still corresponds to lifetimes of several picoseconds. As we will see, these seemingly low Q factors are no barrier to polariton formation.

In Fig. 4.2(b), we add the doped QWs with an ISB resonance at 3 THz and a linewidth 5 % of the resonance frequency ($\gamma = 0.15$ THz). Now we see a clear anti-crossing between the photon mode and the QW mode – a hallmark of polariton formation. So the photon mode is split into two polariton modes, quite similar to the classic exciton-polariton dispersion diagram of Fig. 3.2. Similar to the exciton-polariton diagrams, we have plotted the QW mode as a flat line here. This is because the ISB plasmon mass is expected to be around the effective mass of the conduction band electrons[1]: around $0.067 \cdot m_e$ for GaAs, with m_e being the bare electron mass. On the other hand, from the dispersion curve here, we can see that the photon effective mass is around $3 \times 10^{-8} \cdot m_e$, which is many orders of magnitude lighter. Looking at it another way, these plots only extend to wave vectors of around 10^{-4} nm^{-1}, which is effectively zero from the perspective of the semiconductor dispersions.

In addition to the appearance of splitting, we can see that the magnitude of the splitting is an appreciable fraction of the original transition frequency, even though this structure has not been optimized for maximal strong coupling. These large relative splittings are an especially prominent feature of THz polaritons compared exciton polaritons. However, this comes with a related challenge, which is the large relative linewidth of QW transitions. In Figs. 4.2 (c) and (d), we see what happens when the QW linewidth increases. At a linewidth of 15 %, the splitting is still present, but it begins to blur. At a linewidth of 30 %, the strong coupling is essentially lost. Such broad linewidths are certainly not unrealistic for QWs in the THz, especially when temperatures rise above the cryogenic. For comparison, 30 % is around the QW linewidth used in the room-temperature THz polariton demonstration by Geiser *et al.* [67]. While this was enough to

[1] As a rough justification: in the effective mass approximation, the subband dispersion curves are of the form $\hbar^2k^2/2m^* + E_i$, where m^* is the conduction band electron effective mass, and E_i is the subband minimum. If we assume the effective masses of all subbands are the same, then an ISB transition from subband $i \to j$ with wave vector q will have energy $\hbar^2((q+k)^2 - k^2)/2m^* + E_j - E_i$. When viewed as a function of the transition wave vector, q, we can see that this is a parabola with the exact same curvature as the original subbands. This is not a perfectly rigorous justification, of course, but it demonstrates why we expect the effective mass of the ISB transition (strictly, ISB plasmon) to be similar to that of the underlying electrons.

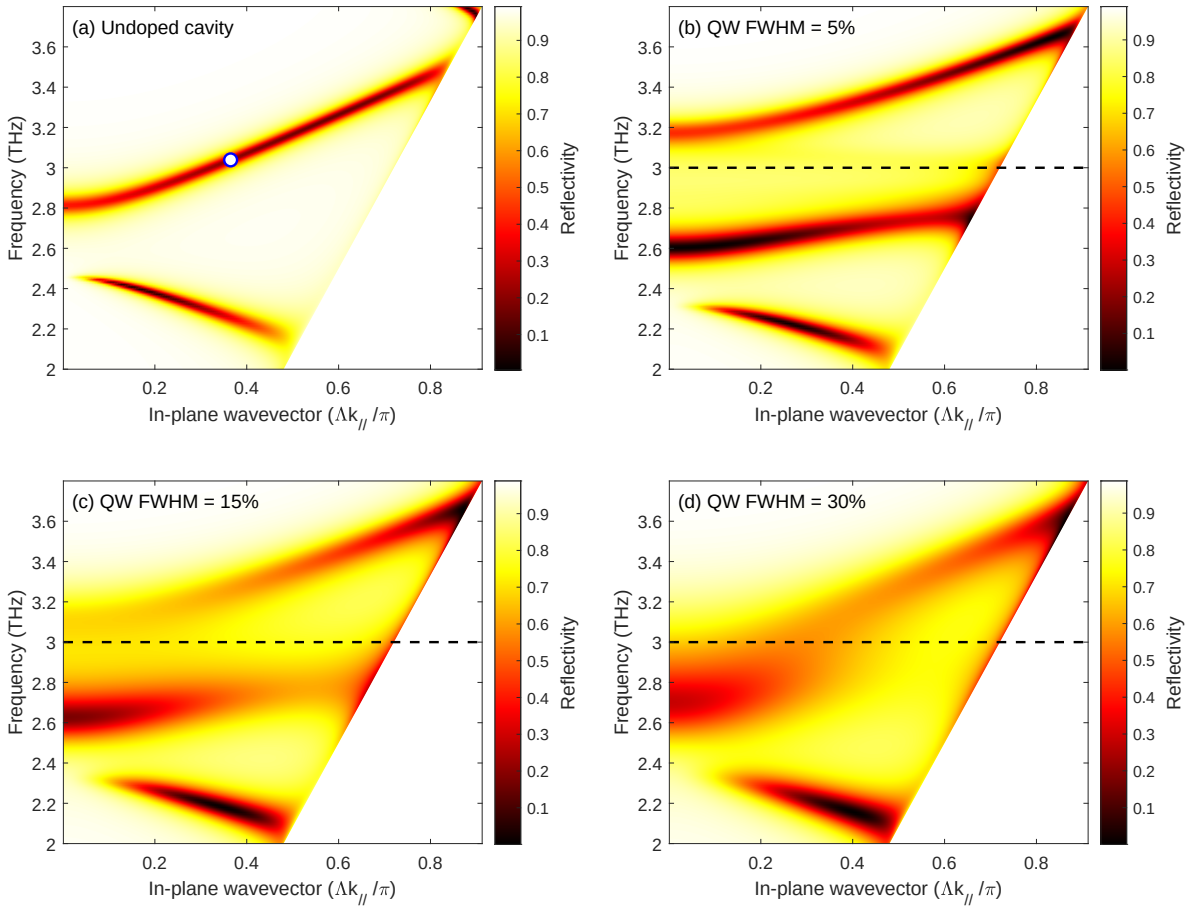

Figure 4.2: Plots of simulated reflectivity of the MIM cavity described in the text as a function of frequency and in-plane wave vector of the incoming light. (In-plane being perpendicular to the growth direction.) The in-plane wave vector is given in units of the grating's Brillouin zone, π/Λ, where $\Lambda = 36\,\mu\text{m}$ is the grating period. Note that there is a forbidden zone in the bottom right corner of each plot, outside the light cone: external photons cannot exist with $\left|k_{\|}\right| > \omega/c$ regardless of the incidence angle θ. **Panel (a):** The bare cavity dispersion, with the QW doping set to zero (or, $\omega_P = 0$). The mode corresponding to the blue and white circle is plotted in Fig. 4.3. **Panels (b)–(d):** The reflectivity with the doped QWs present, for different linewidths, γ, of the QW transition. The linewidth is measured as a percentage of the centre absorption frequency, ω_0. The ISB absorption frequency is plotted as a dashed black line.

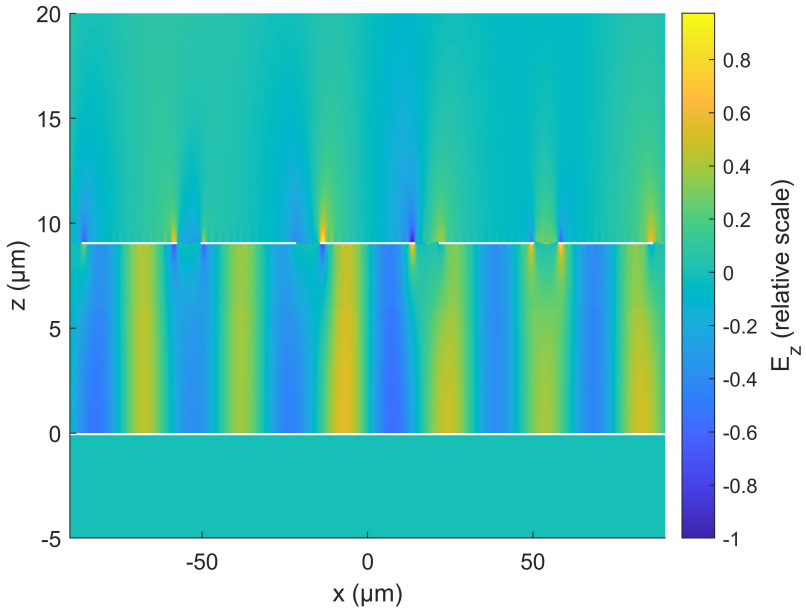

Figure 4.3: The growth-direction component of the electric field **E** for the 3.04 THz mode of the bare MIM cavity at $\theta = 30°$ (corresponding to the blue and white circle in Fig. 4.2). z is the growth direction, and x is the in-plane position perpendicular to the strips. The white regions are the gold layers that form the cavity (the top strips are oriented in the y direction, into the page). Note that the electric field is Bloch-periodic, meaning that it satisfies $\mathbf{E}(x + \Lambda) = e^{ik_{\parallel}\Lambda}\mathbf{E}(x)$. (We are assuming that a large number of strips is present, forming a photonic crystal.)

demonstrate strong coupling in that particular cavity, we see here that the requirements become more stringent when we move towards the dispersive cavities needed for devices.

Of course, the exact numbers will depend on the particulars of the cavity design. There are many parameters that can be adjusted to tune the cavity for a particular problem: e.g., changing the grating periodicity and duty cycle or the active region thickness. This particular cavity is well-tuned for the scattering studies discussed in the next section: the entire dispersion curve is visible on the reflectivity diagram, which means that polaritons at all $k_{\|}$ values (at least, above the light cone) are well-coupled to external cavity modes. This is important for experiments where one is interested in pumping or probing polaritons all along the dispersion curve. On the other hand, if one were designing an electrically-injected laser, say, where only the $k_{\|} = 0$ mode needs to couple strongly to the external world, one may prefer to reduce the contrast at higher $k_{\|}$ values to obtain a steeper dispersion curve.

So it is certainly important to tune the cavity parameters to a particular problem. Nevertheless, this example is broadly illustrative of the requirements for THz active regions in $Al_x Ga_{1-x} As$ devices. For a wide range of cavity parameters, active region linewidths around 15 % or better will be necessary for THz MIM structures aimed at polariton scattering. The higher-quality the active regions are, the more flexibility we have in the overall structure design.

4.4 Stimulated scattering

Strong coupling alone is not enough to make a device with polaritons. For lasing in particular, one of the key additional ingredients is stimulated scattering. To build up a condensate population in the ground state, there needs to be some mechanism which scatters polaritons to the ground state quickly enough to overcome losses. In the case of exciton-polaritons, one of the common scattering mechanisms employed has been phonon-polariton scattering, as polaritons inherit a phonon interaction from their exciton fraction. In the THz and IR, this becomes more difficult. Acoustic phonon scattering is very slow in this regime, although an early proposal for intersubband polariton lasers suggested the use of LO-phonon scattering for the purpose of lasing [58]. There, LO-phonon scattering rates are shown to be on the order of 0.1 ps without including the enhancement due to stimulated scattering. This compares well with polariton lifetimes, which are expected to be on the order of 1–2 ps based on the linewidths of MIM cavities and quantum wells [59].

LO-phonon scattering with ISB polaritons has even been observed experimentally in the mid-IR [84], but unfortunately it presents challenges in the THz. In GaAs, for example, the LO-phonon frequency is around $f_{LO} \approx 8.7\,\text{THz}$, which is far too large compared to the polariton dispersion if one is targeting a minimum around, say, 3 THz. In other words, THz ISB polaritons do not have enough energy to create LO-phonons. One would need to find a more indirect method – for example, creating a higher-energy "dark" plasmon mode which sits at an energy equal to one LO-phonon above the lower-polariton dispersion minimum. Such an approach would likely require sophisticated multi-QW designs analogous to those of THz quantum cascade lasers (QCLs).

A more interesting alternative is polariton-polariton scattering, which is ultimately inherited from electron-electron interactions in the QW active region. Unfortunately the theory of such interactions remains sparse in the context of ISB plasmons coupled to photons. Calculations at the mean-field level have suggested that polariton-polariton scattering rates could be reasonably fast, with interaction energies in the meV range [85]. Still, much work needs to be done here to understand these scattering mechanisms.

More recent work has studied the dynamical modelling of ISB polaritons in MIM geometries using a generalized Gross-Pitaevskii approach [86], similar to an approach used in exciton-polariton condensates

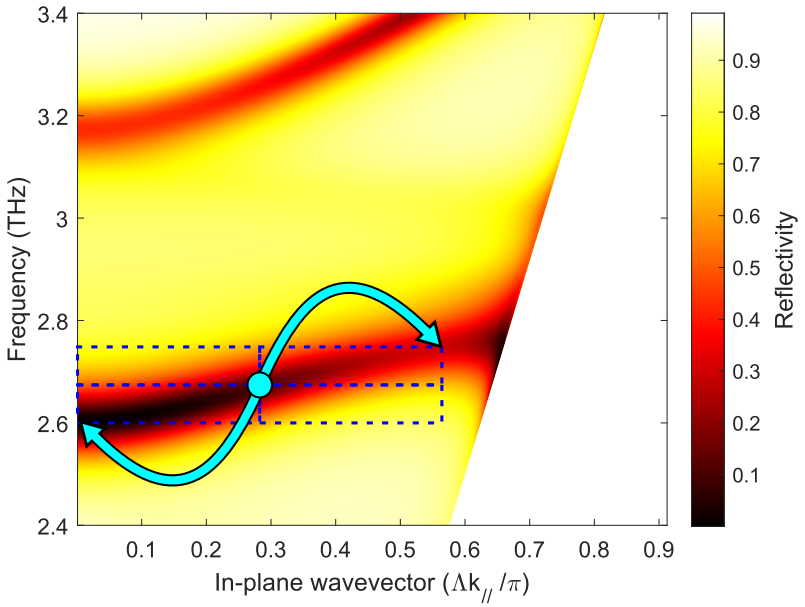

Figure 4.4: A repeat of Fig. 4.2(b), with a possible polariton-polariton scattering interaction shown. Two polaritons at the same energy and momentum scatter off each other to create one low energy and one high energy polariton. Energy and momentum are both conserved, as shown by the dashed rectangles.

[31]. It is shown that several useful non-linear optical phenomena are possible, related to polariton lasing. This model allows for predictions compared to experiment, but there are still interaction coefficients which remain unknown: particularly that of polariton-polariton scattering.

This presents an opportunity which is one of the driving motivations of the later chapters of this thesis. If active regions of sufficiently high quality can be used to make dispersive polariton modes (like those shown in Fig. 4.2 (b)), polariton scattering could be studied experimentally. The results, compared to theory, could provide important insights towards the development of THz intersubband devices.

An example of such a scheme, which is similar to early experiments done on exciton-polaritons [39], is shown in Fig. 4.4. Thanks to the unique dispersion curve of polaritons, there exists a "magic angle" at which two polaritons with the same wave vector, k_m can scatter off each other such that one is kicked into the ground state at $k_{||} = 0$, and the other is kicked into a higher wave vector $2k_m$.

This mechanism can be readily studied by pump-probe experiments. A strong pump beam excites the sample at the magic angle to generate polaritons. Some of these polaritons scatter off each other, creating polaritons at $k_{||} = 0$ and $2k_m$, which can decay to emit photons at angles different from the pump beam. Moreover, if a probe signal is applied at angles corresponding to $k_{||} = 0$ or $2k_m$, it will create polaritons at these wave vectors which will amplify the scattering process, thanks to their bosonic nature. In the THz, pumping with a THz QCL and probing with time-domain pulses is already achievable, meaning that such scattering studies could be performed with an appropriate combination of active region and MIM cavity.

4.5 Pumping to create polaritons

Improved understanding of polariton scattering processes is perhaps one of the most important challenges in THz ISB polaritonics, as it is critical to device operation, and there appears to be a clear line of attack. Nevertheless, it is worth discussing one of the other challenges that will crop up on the road towards polariton lasing: pumping. There are essentially two ways to create a polariton: either through its matter fraction or through its photon fraction. This is because polaritons tend to inherit behaviours from their constituent particles. To a decent approximation, if one has a method for creating the matter part of a polariton (exciton, intersubband plasmon, etc.) or a method for creating the photon part, then one can use the same method to create polaritons.

There is a relatively simple argument for this. In the rotating wave approximation, the polariton creation operators have the form

$$p_{\pm}^{\dagger} = \alpha_{\pm} a^{\dagger} + \beta_{\pm} b^{\dagger}, \tag{4.4}$$

where ± denotes upper or lower polaritons, and $a^{\dagger}$ and $b^{\dagger}$ are the cavity photon and matter excitation creation operators respectively. This relationship can also be inverted:

$$a^{\dagger} = \alpha'_{-} p_{-}^{\dagger} + \alpha'_{+} p_{+}^{\dagger} \tag{4.5}$$

$$b^{\dagger} = \beta'_{-} p_{-}^{\dagger} + \beta'_{+} p_{+}^{\dagger} \tag{4.6}$$

and it turns out that $\left|\alpha'_{\pm}\right| = |\alpha_{\pm}|$ and $\left|\beta'_{\pm}\right| = |\beta\pm|$. In other words, photons and matter excitations can be written as a sum of upper and lower polaritons. This allows us to write interactions involving photons or matter excitations as interactions involving polaritons. For example, suppose cavity photons interact with an external pump light through an interaction Hamiltonian [23]

$$H_{\text{pump}} = g\left(K_{\text{pump}} a^{\dagger} + K_{\text{pump}}^{*} a\right) \tag{4.7}$$

Then, using the expression above, we immediately obtain

$$H_{\text{pump}} = \sum_{\pm} \alpha'_{\pm} g\left(K_{\text{pump}} p_{\pm}^{\dagger} + K_{\text{pump}}^{*} p_{\pm}\right) \tag{4.8}$$

So polaritons obey the exact same interaction, just with the interaction energy multiplied by the coefficient $\alpha'_{\pm}$, which has the same magnitude as the polariton's light fraction $|alpha_{\pm}|$. The more or less "photon-like" the polariton is, the more it will inherit interactions from the photons.

It's not hard to see how this argument would generalize to any interaction written in the second-quantized form. Then, if the interaction is perturbative and can be treated with Fermi's golden rule, scattering rates will be preserved with only two adjustments:

1. Scattering rates are multiplied by the appropriate Hopfield coefficients $|\alpha_{\pm}|, |\beta_{\pm}|$.
2. Energy conservation is based on the polariton energy rather than the constituent particle energy.

Strictly speaking, this picture breaks down in the ultra-strong coupling regime since the rotating wave approximation is no longer valid [9]. In this case the interactions will pick up additional off-resonant terms. However, in cases where these terms can be neglected, this picture should still serve as a useful guide.

4.5.1 Electrical injection and the dark plasmon problem

For the creation of compact devices, the most desirable method for pumping a polariton laser would be electrical injection, in which polaritons are created by an external voltage or current. In some sense electrical injection is relatively easy for intersubband polaritons. The top and bottom of a MIM cavity can naturally be used as contacts to inject electrons and create polaritons through their intersubband plasmon fraction.

However, there is a major challenge with electrical injection, which is related to so-called "bright" and "dark" plasmon modes. When one considers all the possible transitions of an electron from one subband to another (including superpositions of single-electron transitions), it turns out that there are certain special transitions that couple most intensely with light: the "bright" states [78]. Conversely, the large number of transitions which are orthogonal to these bright states will not couple with light at all. Unfortunately, the number of bright states tends to be a constant number, while the number of dark states scales with the number N of electrons in the cavity. This is a problem, because the simplest implementations of electrical injection will create arbitrary excitations with no preference towards the bright ones – in this case, the probability of injecting a bright plasmon is of order $1/N$ [60, 87]. In fact, this is not only a problem for electrical injection, but could be a problem for any pumping scheme which creates intersubband polaritons through their matter fraction.

This is predicted to be a major challenge in electrical injection schemes [87], and has already been seen in experiments [60, 64]. To get around this, one would have to design an electrical pumping scheme which selectively excites the bright modes. This may be possible through energy selectivity: the energy of the bright modes is shifted away from the dark modes because of the interactions that lead to plasmon and polariton formation. If the polariton splitting pushes their energy far enough away from the dark mode energy, it may be possible to selectively excite polaritons through their bright plasmon part without simultaneously exciting dark plasmons. In practice this might be achievable with a filter QW in a quantum cascade structure, although this requires more complex structures which could be especially challenging when approaching room temperature. The larger the coupling strength (and thus the Rabi splitting), the more feasible this may become.

4.5.2 Optical pumping

An easier way to avoid the dark plasmon problem is through THz optical pumping. This works because a THz pump light creates intersubband polaritons through their cavity photon part rather than their plasmon part. The downside, of course, is that we are trying to produce THz light in the first place and good THz pump lights are difficult to come by. Still, while this would not be practical in a final device, it could be an important method for intermediate demonstrations and studies.

Another alternative would be to try interband optical pumping with much more readily available near-IR sources. The idea would be to create polaritons by exciting electrons from the valence band into the conduction band. However, this is fundamentally different from optical pumping with THz light, because it creates intersubband polaritons via their plasmon fraction and not their THz photon fraction. Thus, the dark plasmon problem becomes a concern again. To avoid this, again the most obvious approach would be to rely on the small energy difference between bright and dark states to selectively excite the bright states. This means that a very narrow-band ($\ll 1\,\mathrm{THz}$) optical source would need to be used. Such sources do exist, and could potentially be useful for pumping in pump-probe experiments. Optical pumping – either THz or near-IR – would not likely be useful for practical devices in the end, but will likely play a major role in the intermediate study and development of such devices.

4.6 Final comments

In this chapter, we have discussed some design aspects and challenges related to the development of ISB polaritons lasers in the THz. In particular, we have identified scattering mechanisms as a key area that still needs study. Pump-probe experiments with MIM cavity-based geometries could help to elucidate some remaining unknowns. However, it will be important to have THz QWs with high-quality resonances, because the cavity geometries required for scattering and lasing place more stringent requirements on the active regions. In the remaining chapters, we will therefore turn our attention to the modelling, design, and growth of active regions for THz polariton devices. While our motivation is the road towards THz polariton lasing, such developments could certainly be of interest to the broader THz community. Even within THz polaritonics, there is much more potential for devices besides lasers [88, 89], and many theoretical questions still remain to be answered.

Chapter 5

Modelling intersubband absorption in quantum wells

In the previous chapter, we looked at some design aspects of terahertz (THz) intersubband (ISB) polariton devices with an eye towards scattering and lasing. We saw that active regions with certain absorption characteristics can, when placed in a metal-insulator-metal (MIM) cavity, give rise to THz polaritons with dispersions appropriate for scattering. However, until now, we have been largely ignoring the underlying details of the active region, simply treating it as a "black box" absorber with a Załużny-Nalewajko-like permittivity tensor, Eq. (4.1). We will now turn our attention to those underlying details, and explore how to model the quantum well (QW) active region absorption – this will be essential if we want to design active regions that with a particular THz absorption frequency and absorption strength.

Such design challenges have been well-studied [90] in the case of lightly-doped QWs at low temperature – specifically, QWs in which only the bottom subband is initially occupied before light is absorbed. Less well-known, however, is the behaviour of QWs with several occupied subbands. This regime is often encountered in the design of ISB polaritonics, since relatively heavy doping levels are needed to achieve strong coupling. In this regime, several different ISB transitions can merge into combined modes, quantitatively and qualitatively changing the behaviour of the overall QW absorption. It is therefore critical to be able to accurately capture such effects when designing an active region.

5.1 Schrödinger-Poisson modelling

At the core of our QW modelling is the 1D Schrödinger equation. By definition, a QW structure is homogeneous in two dimensions perpendicular to the growth axis, and heterogeneous along the growth axis. To count as a "quantum" well, this heterogeneity should be on a small enough scale (usually nanometres) to quantize the electron energy levels. However, since the confinement is only in one dimension, the quantization is only partial: the conduction band of the bulk crystal is broken into a set of *subbands*. Mathematically speaking, for a bulk crystal, the energy dispersion of electrons in the conduction band takes the form (in the effective mass approximation)

$$E(\mathbf{k}) = \frac{\hbar^2 |\mathbf{k}|^2}{2m^*} \tag{5.1}$$

where $\mathbf{k} = \left(k_x, k_y, k_z\right)$ is the wave vector and m^* is the effective mass. In a QW, on the other hand, the dispersion is no longer continuous in the z (growth) direction, and we obtain a set of subband dispersions, the ith dispersion having the form

$$E_i(\mathbf{k}_\parallel) = E_{i,0} + \frac{\hbar^2 \left|\mathbf{k}_\parallel\right|^2}{2m^*} \tag{5.2}$$

where $\mathbf{k}_\parallel = (k_x, k_y)$ is the in-plane part of $\mathbf{k}$, and $E_{i,0}$ is the subband minimum.

In certain approximations, QWs bear a close resemblance to the 1D Schrödinger equation models studied in a typical undergraduate quantum mechanics course. Essentially, in a QW, the conduction band edge as a function of z can be taken as the potential energy term in an effective 1D single-particle Schrödinger equation. Solving for the discrete eigenenergies of that equation gives the subband minima, $E_{i,0}$, and the eigenfunctions give the z-varying part of the envelope wavefunctions in the QW. ("Envelope" meaning that we have separated out the underlying atomic-scale Bloch wavefunctions.) So there is a very close analogy between QWs and simple 1D Schrödinger problems, which is useful to keep in mind as an intuitive picture. However, one must remember that this is not a perfect correspondence: the other two dimensions are still present, and QWs come with additional complexities and features that are not present in single-particle 1D Schrödinger systems.

5.1.1 Self-consistent solutions

One such complexity is that we have many electrons present, not just one. So, strictly speaking, this is a many-body system with a many-body wavefunction in $6N$-dimensional configuration space (with N being the number of particles with two spin degrees of freedom). To use the single-particle Schrödinger equation, we must employ the mean-field approximation. That is, we focus on a single electron, and we assume that all the other particles, on average, simply contribute to the electric potential term of that particle's Schrödinger equation. The electric potential V, can be calculated via Poisson's equation $\nabla^2 V = -\rho/\epsilon$, where ρ is the electric charge density from all the other particles and ϵ is the dielectric constant. However, ρ depends on the probability of electrons appearing at a certain location. In the case of a quantum-confined system, these probabilities must be calculated via the eigensolutions to the Schrödinger equation and statistical mechanics.

The result is a self-referential model which is usually difficult to solve analytically. Fortunately, the model is readily solved by numerical methods – one can iteratively solve the Schrödinger and Poisson equations until a solution converges. These Schrödinger-Poisson solvers are the bread and butter of nanostructure modelling, and there are robust code bases that already exist. Throughout this thesis, we will employ the commercial tool nextnano++ [91]. (Further details of the Schrödinger-Poisson approach used by nextnano++ can be found in Ref. [92].)

5.1.2 Some example calculations

Later in this chapter, we will compare the absorption behaviour of two different types of QWs: square and parabolic. We will need more machinery than the self-consistent Schrödinger-Poisson to calculate the absorption, but to get a taste for these systems and for the Schrödinger-Poisson results, we can already calculate their wavefunctions and energy dispersions. We will look at two model systems: square QWs and parabolic quantum wells (PQWs) in $\mathrm{Al}_x\,\mathrm{Ga}_{1-x}\,\mathrm{As}$ with periodic boundaries (effectively simulating an infinite periodic array of wells). Both are designed such that the spacing between their bottom two subbands

is around 12.4 meV at 0 K, which corresponds to a photon frequency of 3.0 THz. The wavefunctions and subband dispersions of these systems in the undoped case are shown in Fig. 5.1, and the details of their structure are provided in the figure caption. These are the results of nextnano++ simulations in the effective mass approximation (plus a correction to the dispersions, which will be discussed in Sec. 5.1.3).

Already, we can see two features that will become important. First, the subbands for the PQW are nearly evenly spaced – analogous to the simple quantum harmonic oscillator – while the square well's subbands get further apart at higher energies. Second, with these simulations at 300 K, we see that there is significant electron occupation of several subbands. Thus, we can expect that several transitions will be active at the same time. This is essentially unavoidable at room temperature in the THz, since at room temperature we have $k_B T \sim 26\,\text{meV}$, which is larger than the small subband spacings required for THz photon emission. As we will see later, the PQW provides a much more robust THz absorption line, particularly above cryogenic temperatures.

We must be careful, however, not to rely too heavily on these simple diagrams – Schrödinger-Poisson is only the starting point for the calculation, and the absorption behaviour can deviate significantly once we include the multisubband plasmon effects discussed in Secs. 5.3 and 5.4. Even the Schrödinger-Poisson picture can vary considerably with doping and temperature – particularly at higher doping levels where the Hartree potential becomes significant. This is exemplified in Fig. 5.2, where the same square QW and PQW system are subjected to a doping level of $3 \times 10^{11}\,\text{cm}^{-2}$. We can see that in both cases the conduction band edge is significantly distorted compared to the undoped case. Note, for example, that the PQW has been reduced to nearly half its original depth thanks to the Hartree potential created by the electrons. As a result, its subbands are much closer together. Thus, looking at this picture, one might expect that the absorption frequency of the PQW would be significantly redshifted at this doping level. However, we will see that this is not actually the case once the multisubband plasmon effects are included: the PQW absorption frequency is largely unaffected by doping. So, while these Schrödinger-Poisson simulations are an important starting point, we must remember that they are not the complete picture.

5.1.3 Subband non-parabolicity

When performing these calculations, there is another feature of semiconductor electrons which differs from their free-space electron counterparts that we must be cognizant of: their dispersion is not exactly quadratic. In the simple dispersion equations of Sec. 5.1, we employed the *effective mass approximation*: we assumed that the electrons in a semiconductor maintain the quadratic energy dispersion of a free particle, just with a modified curvature. Mathematically, this modified curvature is equivalent to a change in the particle's mass, so we simply assign the particle an effective mass to get the correct curvature. However, in general, such an approach only works for $\mathbf{k}$ values close to the subband minimum where the dispersion can be expanded as a second-order Taylor series. At higher $\mathbf{k}$ values, higher-order terms become more prevalent, and the quadratic approximation no longer applies.

For III-V semiconductors, one of the most popular ways to describe the electron and hole dispersions more accurately is the $\mathbf{k} \cdot \mathbf{p}$ method. (An introduction relevant to photonics can be found in Ref. [76].) There are different levels of complexity in the $\mathbf{k} \cdot \mathbf{p}$ method, depending on how many bands are included in the calculation. Nextnano++ allows for an 8-band approach, which includes 3 hole bands and 1 conduction band, each with 2 spin degrees of freedom. This 8-band approach is currently one of the state-of-the-art techniques for modelling semiconductor quantum heterostructures, but it comes with downsides for the applications studied here.

For one thing, the time it takes to perform a simulation increases dramatically compared to the effective mass approximation. Performing Schrödinger-Poisson calculations for a single quantum well in

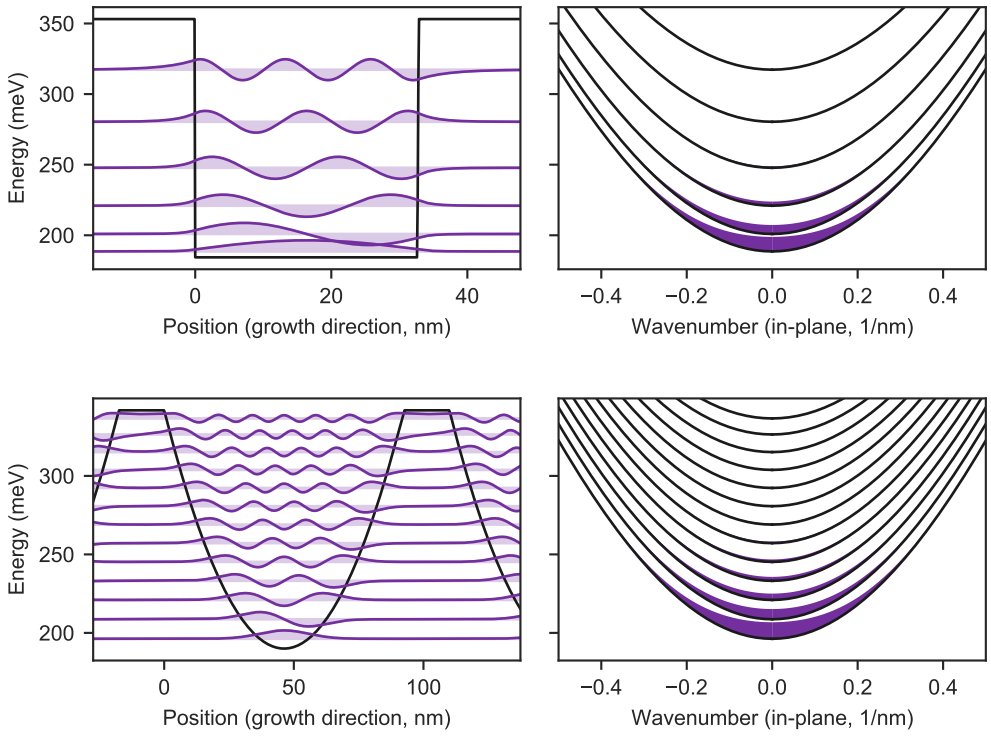

Figure 5.1: Calculated wavefunctions and subband dispersions for two different QW systems. **Top panels**: a square GaAs QW, 32.8 nm wide, with $Al_{0.20}Ga_{0.80}As$ barriers. **Bottom panels**: a PQW in smoothly-graded $Al_xGa_{1-x}As$ with $0.02 \leq x \leq 0.20$. The width of the PQW is 92.5 nm, up to its truncation at an $Al_{0.20}Ga_{0.80}As$ barrier. In both cases, periodic boundary conditions are assumed with a periodicity of 110 nm. Both wells are lightly delta-doped in the centre of the barrier at 1×10^9 cm^{-2}. (At this level, the Poisson effects are negligible.) The calculation is performed at 300 K. **Left panels**: the conduction band edge (black) and wavefunctions (purple), as calculated by Schrödinger-Poisson simulations with nextnano++. The vertical scale of the wavefunctions is arbitrary, but they are offset by their corresponding energies. **Right panels**: the subband dispersions (black) as a function of in-plane wave number, $k_\parallel$. The relative level occupation as a function of $k_\parallel$ is shown for each subband by the height of the purple filled regions. In all cases, energy is measured relative to the Fermi level.

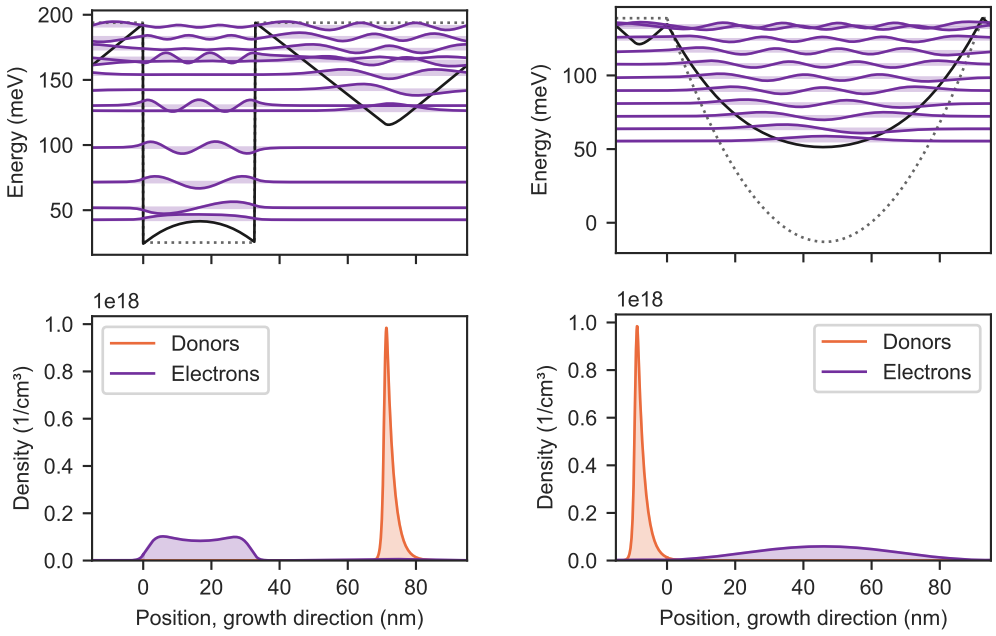

Figure 5.2: Envelope wavefunctions for the same QW systems as Fig. 5.1, but modulation doped at 3×10^{11} cm^{-2} per well. These are calculated self-consistently (i.e., including the Hartree potential) with the Schrödinger-Poisson method at 300 K. **Left panels**: the square QW system. **Right panels**: the PQW system. **Top panels**: the conduction band edge (black, solid) and the wavefunctions (purple). Energy is measured relative to the Fermi level, and wavefunctions are offset according to their subband minima. For reference, the undoped conduction band edge is plotted (grey, dotted). The undoped conduction band edge is shifted so that it lines up with the top of the doped conduction band edge. **Bottom panels**: Electron and Si donor densities. The electron densities are calculated in the self-consistent method. The donor density is a 2D delta density which has been smeared using an asymmetric exponential function to approximate the effect of Si segregation during growth [93].

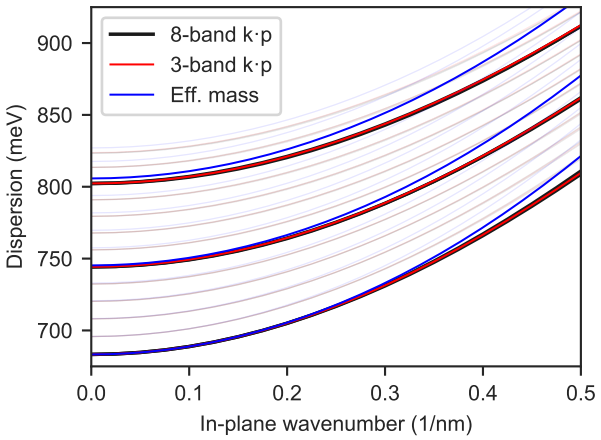

Figure 5.3: Subband dispersions for an undoped PQW with a 3 THz transition frequency, calculated using different methods. Only a few arbitrary subbands are highlighted and the rest are faded for clarity. **Blue**: the subband minima are calculated by nextnano++ in effective mass mode. These are extrapolated into subbands by using a weighted average of the effective mass in the well (which depends on composition) [76]. The weighting function in the average is the probability density $|\psi_n(z)|^2$. **Black**: the entire dispersions are calculated by nextnano++ using its 8-band $\mathbf{k}\cdot\mathbf{p}$ mode. (Note that for this model, the spin degeneracy is lifted, and each subband is actually split into two. However, the dispersions are so close together that the difference cannot be seen in this plot.) **Red**: the dispersions are calculated using the 3-band $\mathbf{k}\cdot\mathbf{p}$ approach described in the text.

the effective mass approximation can take seconds on a standard laptop computer, while a full 8-band calculation could take hours. Such slow calculations can make it much more difficult to test out different device designs.

More critically, though, the mathematical structure of $\mathbf{k}\cdot\mathbf{p}$ is different from that of the effective mass approximation. The envelope wavefunctions are no longer simply functions of space, but they are spinors: including components from each of the bands. In other words, the conduction and hole wavefunctions become "mixed" together when including the $\mathbf{k}\cdot\mathbf{p}$ interactions. Furthermore, the wavefunction spinors are not constant within each subband, but they vary as a function of wave vector.

This change in the mathematical structure of the wavefunctions presents challenges for the models that follow, which have been developed under the effective mass approximation. We have developed a new model in an attempt to tackle these challenges (thanks, primarily, to the theoretical efforts of Dr. Wojciech Pasek) and were in the process of writing up these results at the time of thesis preparation. However, for the purposes of this thesis, we will not attempt to include the 8-band $\mathbf{k}\cdot\mathbf{p}$ wavefunctions into our model.

Still, the effective mass approximation is problematic because it does not provide an accurate calculation of the conduction band dispersions. See Fig. 5.3: not only is the curvature incorrect, leading to a divergence at high $|k_\parallel|$ values, but we can see that the subband minima at $k_\parallel = 0$ calculated by the effective mass model are incorrect as well. These deviations are on the order of several meV, which is problematic, because we are interested in THz ISB transitions, which correspond to energy spacings around 10 meV. So we see that the effective mass approach is not likely to predict the correct ISB absorption frequency in

QWs, particularly when higher subbands and higher $|k_\parallel|$ values are occupied.

Fortunately, while the issue of wavefunctions in 8-band $\mathbf{k}\cdot\mathbf{p}$ is quite difficult to circumvent, it turns out that we can at least attempt to solve the energy dispersion part of the problem by using a mixture of effective mass and 3-band $\mathbf{k}\cdot\mathbf{p}$ modelling. We can start with the effective mass calculations (done with nextnano++ in our case), and then apply a correction to the dispersion relation using $\mathbf{k}\cdot\mathbf{p}$ theory. To obtain this correction, we use the 3-band $\mathbf{k}\cdot\mathbf{p}$ equations, following Warburton *et al.* [94]. There, the subband energy dispersion $E_n(k_\parallel)$ is given implicitly by the Schrödinger equation

$$H\psi_n(z) = E_n(k_\parallel)\psi_n(z), \tag{5.3}$$

with $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian

$$H = -\frac{\hbar^2}{2m^*(z)}k_\parallel^2 - \frac{\partial}{\partial z}\frac{\hbar^2}{2m^*(z)}\frac{\partial}{\partial z} + E_g(z) - \delta(z) \tag{5.4}$$

and effective mass defined by

$$\frac{1}{m^*(z)} = \frac{E_P}{3m_e}\left(\frac{2}{E_n(k_\parallel)+\delta(z)} + \frac{1}{E_n(k_\parallel)+\Delta(z)+\delta(z)}\right) + (1+2F). \tag{5.5}$$

Here m_e is the bare electron mass. $E_g, E_P, F, \Delta, \delta$, and F are the band gap, Kane energy, Kane parameter (as defined in Ref. [95]), spin-orbit splitting, valence band offset respectively. (For our simulations, we use the Fermi level as an energy reference, but this is an arbitrary choice.) We are using here a slightly different definition of the parameters than Warburton *et al.*, to be consistent with Vurgaftman *et al.* [95], who have presented values for these $\mathbf{k}\cdot\mathbf{p}$ parameters for many semiconductor alloys. $\psi_n(z)$ is the envelope wavefunctions associated with the n^{th} subband.

In principle, we could add a Poisson equation and solve this system self-consistently to obtain both the wavefunctions, $\psi_n(z)$ and the dispersions $E_n(k_\parallel)$ from these equations. However, we would again run into the problem that the wavefunctions would depend on $k_\parallel$. Instead, as an approximation, we calculate the wavefunctions $\psi_n(z)$ self-consistently (i.e., including band bending due to doping) in the effective mass method, which gives us a single wavefunction for each subband. If we then take these wavefunctions as given, we can then estimate the 3-band $\mathbf{k}\cdot\mathbf{p}$ dispersion by choosing $E_n(k_\parallel)$ so that it minimizes the overlap error,

$$\text{Error} = \int \left[\psi_n^*(z)H\psi_n(z) - E_n(k_\parallel)|\psi_n(z)|^2\right]\mathrm{d}z\,, \tag{5.6}$$

in the equations above. Thus, we will have wavefunctions that have been calculated in the effective mass approximation, and subband dispersions which have been corrected using the 3-band $\mathbf{k}\cdot\mathbf{p}$ equations.

It is not necessarily obvious that this will give the correct dispersion, since the wavefunctions have only been calculated approximately, however the error turns out to be quite small. (This is not unusual in perturbation or variational-style methods, where energies are often easier to estimate accurately than wavefunctions.) An example of subband dispersions calculated in this way is shown in Fig. 5.3. We can see that the subband dispersions calculated by this model are nearly indistinguishable from those calculated in the 8-band $\mathbf{k}\cdot\mathbf{p}$ approach. Thus, while this approach does not give us the full wavefunction spinors, it at least gives a very good approximation of the subband energy dispersions.

5.2 Multisubband plasmons: introduction

Our primary goal in this chapter is to calculate the effective permittivity of THz QWs so that we can understand their absorption behaviour and their interaction with a photon cavity. First, let us recall the simple version of QW absorption, which occurs in the limit of low doping and low temperature. Under these conditions, the permittivity tensor can be quite straightforwardly calculated from the Schrödinger-Poisson results, using the Załużny-Nalewajko approach [61] described in Ch. 4. In this simple model, light is absorbed when electrons jump from one subband into a higher one, and we recall that the absorption frequency, ω_0, corresponds to the energy spacing, $\hbar\omega_0$, between the subbands. The strength of the absorption is proportional to the square of the plasma frequency, ω_P^2 (defined in Eq. 4.2), which in turn is proportional to the doping density.

Our simple picture requires adjustments, however, as the doping level is increased. The first effect to appear is the depolarization shift [96], which arises when the plasma frequency becomes non-negligible compared to the bare ISB frequency. The absorption frequency is blue-shifted, taking on a value

$$\tilde{\omega}_0 = \sqrt{\omega_0^2 + \omega_P^2} \tag{5.7}$$

which is higher than the ISB energy spacing predicted by Schrödinger-Poisson. The correct dielectric tensor can still be obtained through the Załużny-Nalewajko model, but one must use $\tilde{\omega}_0$ rather than ω_0 in Eq. 4.1. (In fact, this is noted in the Załużny-Nalewajko paper [61].)

As we increase the doping further, more interesting effects appear. If a quantum well is doped heavily enough, the bottom subband will "fill up", and the higher subbands will be occupied, even at 0 K. In this case, we have several ISB transitions which will participate in the absorption (exciting electrons from subbands $1 \to 2, 2 \to 3$, etc.). Importantly, these transitions are not necessarily independent of each other. For example, we could end up with an absorption mode which corresponds to partial excitations of both the $1 \to 2$ transition and the $2 \to 3$ transition at the same time (i.e., a superposition of the two transitions), and it could happen at a frequency which is different from both of the bare transitions. These combined modes are called multisubband plasmon (MSP) modes, and they require more sophisticated modelling to capture correctly.

In the following, we will explore two different approaches to modelling MSP modes: quantum and semiclassical. The quantum approach has the advantage of treating light and matter on the same footing. It also gives a somewhat more intuitive sense of how MSP modes are formed, so we will begin with a sketch of this approach. However, the quantum model is somewhat more limited and unwieldy when making quantitative predictions for the types of problems we will examine here. So, for the actual calculations, we will turn our attention to the semiclassical model.

5.3 Multisubband plasmons: the quantum model

The quantum model we examine here has been largely developed and outlined by Y. Todorov and coworkers [78, 79, 97, 98]. We will not repeat their fully detailed description of the theory, but will attempt to give a more digestible overview of some of the key ideas and results.

As we have seen already, the typical problem geometry comprises a stack of doped semiconductor QWs placed inside a photonic cavity. The quantum theory begins with a second-quantized light-matter

interaction Hamiltonian in the electric dipole gauge [78]:

$$H_{\text{int}} = \int \frac{1}{\epsilon_0 \epsilon_s}\left(-\hat{\mathbf{D}}(\mathbf{r}) \cdot \hat{\mathbf{P}}(\mathbf{r}) + \frac{1}{2}\hat{\mathbf{P}}^2(\mathbf{r})\right) \mathrm{d}^3\mathbf{r} \tag{5.8}$$

where ϵ_s is the static dielectric constant of the semiconductor QW system, and $\hat{\mathbf{D}}, \hat{\mathbf{P}}$ are the quantum operators for the electric displacement field and polarization density, respectively. The choice of dipole gauge comes with advantages in this context – for one, it naturally includes the ultra-strong coupling regime [9] (unlike the standard James-Cummings approach neglecting anti-resonant terms [23]). Furthermore, the dipole-gauge Hamiltonian is quite naturally divided into two components: the light-plasmon coupling part, H_{l-p}, which corresponds to the $\hat{\mathbf{D}} \cdot \hat{\mathbf{P}}$ term, and the plasmon interaction part, H_p, which corresponds to the $\hat{\mathbf{P}}^2$ term.

To deal with the plasmon part of the Hamiltonian, H_p, we can begin with Schrödinger-Poisson simulations of the QW region, like the approach discussed in previous sections. We then consider transitions from the i^{th} to the j^{th} subband – for brevity, we label the transition as $\alpha \equiv i \to j$. If we assume that all subbands are perfectly quadratic with the same effective mass, then the ISB transition energy $\hbar\omega_\alpha = E_j - E_i$ is independent of $k_\parallel$, depending only on the subband minima E_i, E_j. In this case, we can define an "ISB transition" quasiparticle with creation operator $b_\alpha^\dagger$, which corresponds to the annihilation of an electron in subband i and the creation of an electron in subband j.[1] The non-interacting matter part of the Hamiltonian can then be written as

$$H_e = \sum_\alpha \hbar\omega_\alpha b_\alpha^\dagger b_\alpha. \tag{5.9}$$

I.e., in the absence of interactions, the energy of the system relative to the ground state is simply a sum of the number of excitations, multiplied by their transition energy.

However, there is also an interacting part of the Hamiltonian, H_p, which can be written in terms of these ISB transition quasiparticles as [97]

$$H_p = \frac{e^2}{2\epsilon_0\epsilon_s} \sum_{\alpha,\beta} S_{\alpha,\beta} \sqrt{\Delta N_\alpha \Delta N_\beta}\left(b_\alpha^\dagger + b_\alpha\right)\left(b_\beta^\dagger + b_\beta\right), \tag{5.10}$$

where $\Delta N_\alpha = N_i - N_j$ is the difference in 2D occupation between the subbands, and $S_{\alpha,\beta}$ is a coupling constant that can be calculated from the Schrödinger-Poisson wavefunctions and energies. So we see that there is an interaction between different ISB transitions, which is strongest when it involves transitions from a heavily-occupied subband into a lightly-occupied subband.

We have skipped over the derivation of this expression, but the important thing to see is that the $\hat{\mathbf{P}}^2$ part of the interaction Hamiltonian gives rise to interactions between different ISB transitions. As we have noted previously, when multiple subbands are heavily occupied, their transitions will not behave independently of each other. This interaction is exactly what gives rise to MSP quasiparticles. They arise from the diagonalization of the full interacting matter Hamiltonian, $H_{e-p} \equiv H_e + H_p$.

It is convenient to do this diagonalization in two stages, because we will see that the depolarization shift naturally arises along the way. So, first, we introduce quasiparticles called ISB plasmons, whose creation operators $p_\alpha^\dagger = u_\alpha b_\alpha^\dagger + v_\alpha b_\alpha$ are a linear combination of the ISB transition creation and annihilation

[1] Strictly, the relevant creation operator will actually correspond to a superposition of single-electron transitions at different $k_\parallel$ values, known as the "bright" excitation [78].

operators $b_\alpha^\dagger, b_\alpha$. Using a Bogoliubov-style diagonalization procedure, it is possible to find coefficients u_α, v_α that eliminate the $\alpha = \beta$ part of H_p: i.e., the self-interaction of a single transition $\alpha = i \to j$. What we find is that the Hamiltonian can be written as

$$H_{e-p} = \sum_\alpha \hbar\tilde{\omega}_\alpha p_\alpha^\dagger p_\alpha + \frac{\hbar}{2}\sum_{\alpha\neq\beta}\Xi_{\alpha,\beta}\left(p_\alpha^\dagger + p_\alpha\right)\left(p_\beta^\dagger + p_\beta\right) \tag{5.11}$$

where $\Xi_{\alpha,\beta}$ is a renormalized interaction energy (see Ref. [97]). More notably, the energy of this new ISB plasmon quasiparticle, $\hbar\tilde{\omega}_\alpha$, turns out to be given by $\tilde{\omega}_\alpha = \sqrt{\omega_\alpha^2 + \omega_P^2}$, where $\omega_{P,\alpha}$ is the ISB plasma frequency, given by

$$\omega_{P,\alpha}^2 = \frac{2e^2\Delta N_\alpha \omega_\alpha}{\hbar\epsilon_0\epsilon_s} S_{\alpha\alpha}. \tag{5.12}$$

In other words, in this quantum picture, the depolarization shift of quantum well transitions comes from the self-interaction of a single ISB transition.

However, the depolarization shift is only the first step – we still have remaining interaction terms between different transitions, which cannot be neglected when there is more than one subband occupied. These are, again, dealt with using a Bogoliubov-type transformation. We define *multisubband* plasmon creation operators which are a weighted sum of all the possible $\alpha \equiv i \to j$ transitions:

$$P_n^\dagger = \sum_\alpha a_{n,\alpha} p_\alpha^\dagger + b_{n,\alpha} p_\alpha^\dagger \tag{5.13}$$

with coefficients $a_{n,\alpha}, b_{n,\alpha}$. Ultimately, our goal is to simplify our interacting Hamiltonian to the form of a non-interacting Hamiltonian:

$$H_{e-p} = \sum_n \hbar\Omega_n P_n^\dagger P_n \tag{5.14}$$

As a necessary condition of this, we must satisfy the commutation relation

$$[H_{e-p}, P_n^\dagger] = \hbar\Omega_n P_n^\dagger \tag{5.15}$$

Plugging in Eqs. (5.11) and (5.13) and using the standard creation/annihilation operator commutation relations, it is possible to show that this commutation relation is equivalent to an eigenvalue problem: the eigenvectors correspond to the coefficients $\left[a_{n,1}, a_{n,2}, \ldots, b_{n,1}, b_{n,2}\ldots\right]$, and the eigenvalues correspond to the energies $\hbar\Omega_n$. Numerically, the matrix to be diagonalized can be calculated from the $\tilde{\omega}_\alpha$'s and $\Xi_{\alpha,\beta}$'s [99].

It can be easy to get lost in the mathematical details, but the key point here is that the $\hat{\mathbf{P}}^2$ part of the interaction Hamiltonian leads to an interaction between different ISB transitions. As a result of this interaction, we find that the "true" material excitations of the quantum well system (i.e. the eigenvectors of the Hamiltonian) are MSP modes, which are *superpositions* of different ISB transitions. The MSP modes come with their own frequencies Ω_n, distinct from the ISB transition frequencies ω_α that were calculated by Schrödinger-Poisson. Furthermore, we see that the well-known depolarization shift arises from the exact same interaction in cases where we can neglect the interaction between different transitions $\alpha \neq \beta$.

Now, the final aspect of the quantum theory is the light-matter coupling part of the Hamiltonian. After performing the plasmon diagonalization, the entire light-matter Hamiltonian can be written as [84]

$$H = \sum_n \hbar\Omega_n P_n^\dagger P_n + \hbar\omega_c\left(a^\dagger a + \frac{1}{2}\right) + i\sum_n \frac{\hbar\Omega_{P,n}}{2}\sqrt{F_w\frac{\omega_c}{\Omega_n}}\left(a^\dagger + a\right)\left(P_n^\dagger + P_n\right) \tag{5.16}$$

where ω_c and $a^\dagger$ are the frequency and creation operator for a photon mode, respectively. The quantities $\Omega_{P,n}$ and F_w are effective plasma frequencies and overlap factors for the MSP modes: these are the quantities which control the interaction strength between light and a particular photon mode. They are calculated by writing down the light-matter part of the Hamiltonian, which is known in terms of $\left(a^\dagger + a\right)\left(b_n^\dagger + b_n\right)$ [78], and then re-writing it in terms of $P_n^\dagger, P_n$ using (5.13).

The effective plasma frequency $\Omega_{P,n}$ turns out to be a linear combination of the bare ISB plasma frequencies $\omega_{P,\alpha}$, with weightings related to the coefficients $a_{n,\alpha}, b_{n,\alpha}$. This can have somewhat counterintuitive results, however. If we have a heavily doped QW with several occupied subbands, we might expect it to have several absorption peaks – at least one corresponding to each filled subbands. However, instead we often find that the plasma interactions cause the vast majority of the light-matter interaction strength to concentrate into a single MSP mode. That is, after diagonalizing the plasma interaction Hamiltonian, we find that there is one mode $n = n_0$ with a large effective plasma frequency while the others, $n \neq n_0$ are effectively zero. The result is that heavily doped quantum wells tend to have a single strong absorption peak, which remains even at room temperature [84]. This qualitative change in the absorption behaviour has been exploited to achieve ultra-strong coupling in the mid-infrared (IR) [74]; however, we will see later that things are more delicate in the THz.

To study these effects further, though, we need to be able to calculate the effective permittivity (and thus, the absorption) for a given system. Unfortunately, the quantum model presents challenges in this respect. While the model does make quantitative predictions of both absorption [97] and an effective dielectric tensor [79] (albeit slightly unwieldy), this model has been explicitly limited to the effective mass approximation, in which subbands are parallel to each other. This is problematic since, as we have seen in Sec. 5.1.3, it is important to correctly account for subband non-parabolicity in the study of THz QWs.

For square QWs, there is a simplification which allows the quantum model to be extended to non-parabolic subbands, and an effective dielectric tensor can be calculated [100]. This approach mirrors much earlier work by Warburton *et al.* who showed an absorption concentration effect in heavily doped InAs/AlSb QWs similar to the MSP mode concentration [94]. In the quantum model, the dielectric tensor calculated in Ref. [100] and similar comes thanks to a simplification of the diagonalization problem when all the ISB transitions have the same effective width [97, 99]. Unfortunately, while this assumption of constant effective length holds for the adjacent $i \to i+1$ transitions in square QWs, it does *not* hold for PQWs or other well shapes in general. Even for non-adjacent transitions in square wells, one must revert to the full diagonalization procedure of the quantum model to obtain the correct results [97].

The quantum method almost certainly could be extended in principle to treat non-parabolic subbands more generally, but it is not obvious how best to tackle this. A naive approach of treating each $k_\parallel$ point as an independent ISB transition quickly leads to an intractably large matrix to diagonalize. Furthermore, it is far from clear how to properly re-couch the theory in terms of the mathematical structure of $\mathbf{k}\cdot\mathbf{p}$ theory. So, instead, we now turn our attention to a semiclassical model which can produce many of the same key results as the quantum model. For our numerical calculations, we will use a semiclassical model based on the work of Alpeggiani and Andreani [101]. We will extend their derivation to allow for asymmetric QWs, and we will also make a first attempt at including subband non-parabolicity.

5.4 Multisubband plasmons: the semiclassical model

5.4.1 The non-local susceptibility tensor

At the core of the semiclassical model is the notion of a non-local susceptibility tensor, which has been used to study the light-matter coupling for both ISB transitions [102, 103] and exciton-polaritons [104]. In a typical dielectric material, we have

$$\mathbf{D} = \epsilon_0(\mathbf{E} + \mathbf{P}) = \epsilon_0(1+\chi)\mathbf{E} = \epsilon_0\epsilon\mathbf{E} \tag{5.17}$$

where $\mathbf{E}$ and $\mathbf{D}$ are the electric and displacement fields, respectively, and χ and ϵ are the electric susceptibility and relative permittivity of the material, respectively. To model QWs, we must complicate this simple relationship by introducing both anisotropy and non-locality. In the in-plane (non-growth) directions, x, y, which do not couple with the quantum well, we maintain a simple local relationship:

$$D_x = \epsilon_0\epsilon_{xx}E_x \tag{5.18}$$

$$D_y = \epsilon_0\epsilon_{yy}E_y \tag{5.19}$$

However, in the growth direction, we allow for a non-local susceptibility, and the constitutive relation is modified to:

$$D_z(z) = \epsilon_0\epsilon_s + \epsilon_0\int \chi_{zz}(z,z')E_z(z')\,\mathrm{d}z' \tag{5.20}$$

Here, ϵ_s is the static permittivity of the semiconductor in the absence of doping. Often, we can assume $\epsilon_s = \epsilon_{xx} = \epsilon_{yy}$, however it may be convenient to allow $\epsilon_{xx}, \epsilon_{yy}$ to have additional frequency dependence to account for in-plane plasma oscillations, similar to Ref. [83].

The non-local susceptibility $\chi_{zz}(z,z')$ is the key component of this relationship, and it can be found in terms of the QW wavefunctions [101, 102]. It can be written in the form:

$$\chi_{zz}(\omega,\mathbf{q};z,z') = \sum_\alpha \chi_\alpha(\omega,\mathbf{q})\xi_\alpha(z)\xi_\alpha(z') \tag{5.21}$$

where $\alpha \equiv n \to n'$ is the index of the ISB transition from subband n to subband n', similar to the last section, and $\mathbf{q}$ is the in-plane wave vector.[2] The ξ_α's are called ISB transition current densities, and they are defined by

$$\xi_\alpha(z) = \frac{\hbar e}{2m^*(z)}[\psi_{n'}(z)\partial_z\psi_n(z) - (\partial_z\psi_{n'}(z))\psi_n(z)], \tag{5.22}$$

where $\psi_n(z)$ are the envelope wavefunctions of the QW (calculated by Schrödinger-Poisson and assumed to be real), and $m^*(z)$ is the effective mass. (In general, the effective mass varies with the composition of the semiconductor, and so it will depend on z.) The other part, $\chi_\alpha(\omega)$, is called the "single-particle susceptibility" and is given by

$$\chi_\alpha(\omega) = -\frac{1}{\omega^2}\frac{4}{\epsilon_0\hbar\omega_\alpha S}\sum_{\mathbf{k}}\Delta f_\alpha(\mathbf{k})\left[1 + \frac{\omega_\alpha(\mathbf{k})\omega_\alpha(0)}{\omega^2 - \omega_\alpha(\mathbf{k})^2}\right] \tag{5.23}$$

[2] It is implied here that we have Fourier-transformed Maxwell's equations in time and in the in-plane spatial variables. See Appendix A for more details.

where $\omega_\alpha(\mathbf{k}) = \omega_{n'}(\mathbf{k}) - \omega_n(\mathbf{k})$ is the transition energy from subband $n \to n'$ at in-plane wave vector $\mathbf{k}$, and $\Delta f_\alpha(\mathbf{k}) = f_n(\mathbf{k}) - f_{n'}(\mathbf{k})$ is the difference in Fermi occupation probability. S is the in-plane surface area.

If we assume that the subbands are two-fold spin-degenerate and isotropic in-plane, then we can do the standard transformation from sum to integral

$$\frac{1}{S}\sum_{\mathbf{k}} \to 2\int_0^\infty \int_0^{2\pi} \frac{k\,\mathrm{d}k\,\mathrm{d}\theta}{(2\pi)^2}, \tag{5.24}$$

and we obtain

$$\chi_\alpha(\omega) = -\frac{1}{\omega^2}\frac{4}{\epsilon_0 \hbar \omega_\alpha \pi}\int_0^\infty \Delta f_\alpha(k)\left[1 + \frac{\omega_\alpha(k)\omega_\alpha(0)}{\omega^2 - \omega_\alpha(k)^2}\right]k\,\mathrm{d}k \tag{5.25}$$

Strictly speaking, one should be more careful about introducing the subband non-parabolicity (see notes in Ref. [101]), but this approach at least allows us to get closer to the correct transition energies. Work on a technique that properly incorporates the $\mathbf{k}\cdot\mathbf{p}$ framework into the definitions of ξ_α and χ_α is still underway at the time of this writing.

In principle, we have everything we need now to solve electromagnetic problems involving heavily-doped quantum wells. The non-local susceptibility (Eq. (5.21)) for the QW stack can be calculated from Schrödinger-Poisson simulation results, and then it can be plugged into Maxwell's equations via the constitutive relation, Eq. (5.20). However, solving Maxwell's equations with this sort of integral relationship is a relatively uncommon task, and it cannot be done with a standard electromagnetic solver.

One approach, then, is to develop a custom electromagnetic solver that incorporates the non-local susceptibility: Alpeggiani and Andreani [101] accomplish this by deriving the transfer matrix for a non-local QW layer. In the long-wavelength approximation, this approach can be simplified even further to give an effective *local* permittivity for the QW stack (similar to that of Załużny, but incorporating the multi-subband interactions). In the next sections, we will demonstrate how to solve the electromagnetic equations and define an effective permittivity tensor for the QW stack.

5.4.2 Solving for the electromagnetic fields

In this section, we will describe the solution of the electromagnetic fields with a non-local susceptibility of the form Eq. 5.21. Our method will be slightly more general than that of Alpeggiani and Andreani, as we will not restrict ourselves to QWs with inversion symmetry. The details of this derivation are fairly involved, so in this section we will only highlight the key results. The full details are provided in Appendix A.

We will assume that we have a layered structure, with a QW layer sandwiched inside cavity (such as the MIM cavities discussed in Ch. 4). We are particularly interested in transverse magnetic (TM)-polarized waves: i.e., waves with a non-zero $\mathbf{E}$ component in the z direction (growth direction), since these are the ones that couple to the QWs through the non-local susceptibility. We will assume that the QW region is isotropic in the $x-y$ plane so that $\epsilon_{xx} = \epsilon_{yy} = \epsilon_\parallel$. (NB: the other layers, such as the metallic strips, need not be isotropic, since these layers do not involve non-local susceptibilities, and can be solved through standard methods.)

If we use Eqs. (5.20) and (5.21) along with Maxwell's equations (again, see Appendix A for details), we

can show that the D_z field within the QW region obeys the integro-differential equation

$$\left(\partial_z^2 + k_z^2\right) D_z = -\sum_\alpha \frac{\chi_\alpha(\omega, q_x)}{\epsilon_s} \xi_\alpha(z) \int \xi_\alpha(z') \left(\partial_{z'}^2 + \epsilon_{xx} k_0^2\right) D_z(z')\, \mathrm{d}z' \tag{5.26}$$

where k_z is defined by

$$k_z^2 = \epsilon_\parallel k_0^2 - \frac{\epsilon_\parallel}{\epsilon_s} q_\parallel^2 \tag{5.27}$$

with $k_0^2 = \omega^2/c^2$ and $q_\parallel$ being the in-plane wave number.

As we show in Appendix A, the general solution to this equation has the form

$$D_z(z) = A\cos(k_z z) + B\sin(k_z z) + q_\parallel^2 \sum_\alpha \frac{\chi_\alpha}{\epsilon_s} \left(A F_\alpha^A + B F_\alpha^B\right) \int \xi_\alpha(z') g(z, z')\, \mathrm{d}z' \tag{5.28}$$

where A, B are arbitrary coefficients which will be determined by the boundary conditions of the layer. $g(z, z')$ is the Green's function $\sin(k_z|z - z'|)/2k_z$ The F_α^A, F_α^B coefficients can be calculated through the following matrix relationships:

$$F_\alpha^A = C_\alpha^A - \sum_{\alpha'} \frac{\chi_{\alpha'}}{\epsilon_s} F_{\alpha'}^A \left[I_{\alpha,\alpha'} + q_\parallel^2 D_{\alpha,\alpha'}\right] \tag{5.29}$$

$$F_\alpha^B = C_\alpha^B - \sum_{\alpha'} \frac{\chi_{\alpha'}}{\epsilon_s} F_{\alpha'}^B \left[I_{\alpha,\alpha'} + q_\parallel^2 D_{\alpha,\alpha'}\right] \tag{5.30}$$

with

$$I_{\alpha,\alpha'} = \int \xi_\alpha(z) \xi_{\alpha'}(z)\, \mathrm{d}z \tag{5.31}$$

$$C_\alpha^A = \frac{\epsilon_\parallel}{\epsilon_s} \int \cos(k_z z) \xi_\alpha(z)\, \mathrm{d}z \tag{5.32}$$

$$C_\alpha^B = \frac{\epsilon_\parallel}{\epsilon_s} \int \sin(k_z z) \xi_\alpha(z)\, \mathrm{d}z \tag{5.33}$$

$$D_{\alpha,\alpha'} = -\frac{\epsilon_\parallel}{\epsilon_s} \int \xi_\alpha(z) g(z, z') \xi_{\alpha'}(z')\, \mathrm{d}z\, \mathrm{d}z'\,. \tag{5.34}$$

From these, it is possible to calculate a transfer matrix for the QW region, similar to Ref. [101]. Note that this is a slight generalization of the results in Ref. [101], as we have not assumed symmetric quantum wells, and we have not assumed that $\epsilon_\parallel = \epsilon_s$.

5.4.3 Effective permittivity

Obviously, the general solution in the previous section is somewhat cumbersome, and it has the downside that it cannot be fed into a standard electromagnetic solver. This makes it much more difficult to simulate

entire structures – particularly structures like the MIM geometries we employed in Ch. 4 which are non-uniform in the x–y plane.

For THz QWs, though, we can take advantage of the long-wavelength limit ($q_x, k_0, k_z \approx 0$), since we have well widths $\lesssim 100\,\text{nm}$ and electromagnetic wavelengths $\gtrsim 10\,\mu\text{m}$, even after accounting for the refractive index of the semiconductor. As shown in detail in Appendix A.4, this allows us to define an effective local permittivity for the QW stack. Intuitively, since the electric field is approximately constant on the scale of a single QW, we are able to replace each QW period with an effective medium.

The basic approach (similar to Ref. [61]) is to take an average of the D_z and E_z fields over a single QW period of length L. Then, the effective permittivity (in the z direction) is given by

$$\epsilon_{zz,\text{eff}} = \frac{\langle D_z \rangle}{\epsilon_0 \langle E_z \rangle} \approx \epsilon_s \left[1 - \sum_\alpha \frac{\chi_\alpha}{\epsilon_\parallel} F_\alpha^A \langle \xi_\alpha \rangle \right]^{-1} \tag{5.35}$$

where $\langle \cdot \rangle$ denotes an average over the entire QW region in the z direction:

$$\langle \xi_\alpha \rangle = \frac{1}{L} \int \xi_\alpha(z)\,\mathrm{d}z \tag{5.36}$$

where L is the thickness of the quantum well region. Furthermore, since we are using the long-wavelength limit, we can use that to simplify our equation for the F_α^A's to:

$$F_\alpha^A \approx L \langle \xi_\alpha \rangle - \sum_{\alpha'} \frac{\chi_{\alpha'}}{\epsilon_s} F_{\alpha'}^A I_{\alpha,\alpha'} \tag{5.37}$$

And we recall that

$$I_{\alpha,\alpha'} = \int \xi_\alpha(z) \xi_{\alpha'}(z)\,\mathrm{d}z \tag{5.38}$$

So in the long wavelength limit, we obtain a greatly simplified picture. From the QW transition current densities, $\xi_\alpha(z)$, we solve the matrix equation (A.51) to get the F_α^A coefficients, from which we can calculate the effective permittivity $\epsilon_{zz,\text{eff}}$. This can then be used for the QW region in a standard electromagnetic solver, such as in the simulations we performed in Ch. 4.

Finally, it should be noted that there is actually a much closer correspondence between this semiclassical model and the quantum model than first meets the eye. In Ref. [101] it was shown that this model predicts identical absorption modes to the quantum model, at least in the long-wavelength limit and neglecting non-parabolicity. Loosely, the expression for the effective permittivity, along with the matrix equation (5.37) can be mapped into an eigenvalue problem which mirrors the Bogoliubov diagonalization procedure in the quantum model.

5.5 Absorption coefficient

In the remaining sections, we would like to discuss the absorption properties of different QW designs. One must be careful here, because in general, the absorption spectrum of a QW depends on its surroundings. In other words, we cannot have a well-defined absorption coefficient for a QW active region unless we also specify the structures that surround the active region. In general, the safest approach is to use the effective

permittivity, $\epsilon_{zz,\text{eff}}$, (which *is* well-defined independently of the surroundings) and solve the absorption for a particular situation.

Nevertheless, the absorption coefficient is a commonly-used quantity which is useful for intuitive understanding. To make it well-defined, we will assume that the absorption takes place in a simple "test" context. We imagine a thin layer of QWs, embedded in a semiconductor of infinite extent in both $\pm z$ directions, with dielectric constant ϵ_s. We assume that the dielectric constant of the QWs would also be ϵ_s if they were undoped. In this context, we can define the absorption coefficient as [90, 96]

$$\alpha(\omega) = -\frac{\omega\sqrt{\epsilon_s}}{c}\text{Im}\left[\frac{\epsilon_s}{\epsilon_{zz,\text{eff}}(\omega)}\right] \tag{5.39}$$

This quantity has units of 1/length and can be thought of as the relative power lost per unit length. For a periodic array of QWs with period L_p, this absorption coefficient will – perhaps counterintuitively – depend on the width of the barrier between the wells. Larger barriers mean fewer QWs per unit length, which leads to less absorption per unit length. Thus, it is often more useful to use a dimensionless quantity $L_p\alpha(\omega)$, which corresponds to the fraction of light absorbed per QW. (Again, though, remember that there is an implicit assumption here that QWs are embedded inside an infinite media with dielectric constant ϵ_s. Otherwise, "absorption per QW" is an ill-defined quantity which depends on the surroundings.)

5.6 Absorption linewidths

In the models above, we have thus far neglected absorption linewidths (i.e., broadening), which come in two forms: homogeneous and inhomogeneous. Inhomogeneous broadening arises, for example, when we have a stack of QWs with slightly different absorption frequencies. This could be due to, say, band-bending effects at the boundaries or variations in the growth. Homogeneous broadening comes from non-radiative scattering of the electrons in the well. The most prominent scattering mechanisms are from interfaces, alloy disorder, phonons, and ionized impurities [105]. Interestingly, homogeneous and inhomogeneous broadening mechanisms have different effects on the formation of polaritons – it has been shown that ISB polaritons are somewhat "immune" to inhomogeneous broadening when we have a system of many QWs coupled to the same photon cavity [106].

Inhomogeneous broadening mechanisms can usually be included in the Schrödinger-Poisson model, although it can be laborious and increase the computational time considerably. Homogeneous broadening mechanisms, unfortunately, remain quite difficult to predict from theory beyond order-of-magnitude estimates. This is especially true in the context of MSP modes, which may have altered scattering rates compared to the bare single-particle transitions. As such, in all our calculations we will artificially introduce a broadening parameter, γ, by evaluating the absorption at $\omega \to \omega + i\gamma/2$. This essentially simulates homogeneous broadening with γ being approximately the full width at half maximum (FWHM) of each absorption peak.

5.7 Why parabolic quantum wells?

At last, we are in a position to study the design of QWs for the active region in THz polaritonic devices. Recall that we would like to be able to achieve a sharp resonance in the THz which is stable with temperature. We would also like to be able to engineer the absorption strength via the doping level, as this will aid in designing a final device.

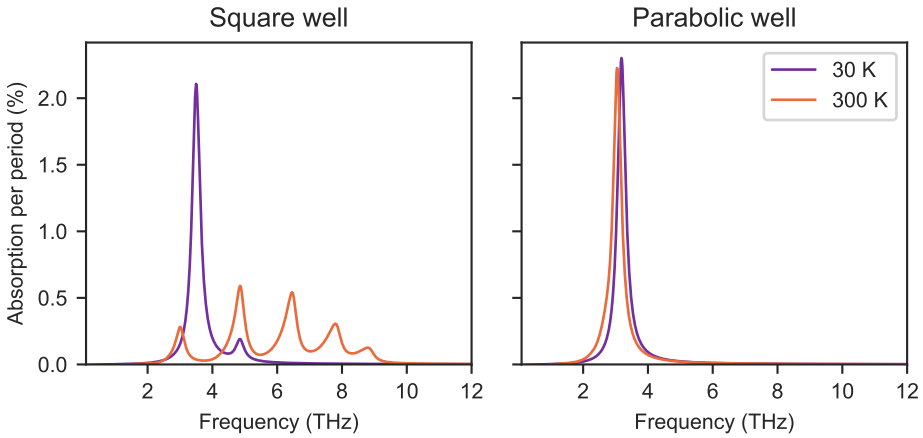

Figure 5.4: Absorption spectra (as defined in Sec. 5.5) for a square QW system (left) versus a PQW system (right). These are the same well systems described in Fig. 5.1, which are designed to absorb around 3 THz at low temperature. In this case, the simulated structures are Si delta doped at 1×10^{11} cm^{-2} per well. We simulate the absorption at 30 K and 300 K. For comparison, we use the same artificial linewidth of $\gamma = 0.3$ THz for all simulations. (In reality, the peaks would likely broaden at 300 K.)

The challenge with square QWs in the THz is immediately apparent if we simulate their absorption at low and high temperature: see Fig. 5.4 (left). While at low temperature we have a peak at 3 THz corresponding to the $1 \rightarrow 2$ ISB transition, at high temperature the absorption splits into several weaker peaks. This is due to the significant occupation of the higher subbands that we saw in our Schrödinger-Poisson simulations in Fig. 5.1. In reality, the problem is even worse than what we see in Fig. 5.4 because the peaks also tend to broaden significantly at higher temperatures due to an increase in mechanisms such as phonon scattering. This broadening diminishes the visibility of the peaks even further in practice. We can also see that this effect appears well below room temperature. If we look at the full evolution of absorption as the temperature rises (Fig. 5.5) we can see that the peak splitting already begins around 50 K. This means that even experiments at liquid nitrogen temperature (77 K) can be challenging.

On the other hand, if we look at the absorption of a PQW (Fig. 5.4, right) we see that it is effectively insensitive to temperature. In reality, we expect that there will be an increase in the linewidth (as we will see in Ch. 8), but the peak position and overall strength of the absorption are quite robust to increases in temperature. Even though there are many transitions present at room temperature, we see a single, clear absorption peak at 3.0 THz. It may seem obvious that this would be the case, given the equally-spaced subbands, but there is actually more going on here than meets the eye. For one thing, due to the non-idealities of the semiconductor system, the subbands are not perfectly equally spaced: the transitions can vary by more than 0.1 THz, even without accounting for subband non-parabolicity. For another thing, at this doping level of 1×10^{11} cm^{-2}, there are enough electrons in the PQW to cause a noticeable distortion to the conduction band edge via their Hartree potential. This potential makes the well a bit shallower, which we would expect to redshift the transition energies. The Schrödinger-Poisson part of these calculations bear this out: while the first ($1 \rightarrow 2$) transition frequency of the undoped system is 3.0 THz, the first transition frequency of the doped system predicted by Schrödinger-Poisson is 2.8 THz. And yet, even in

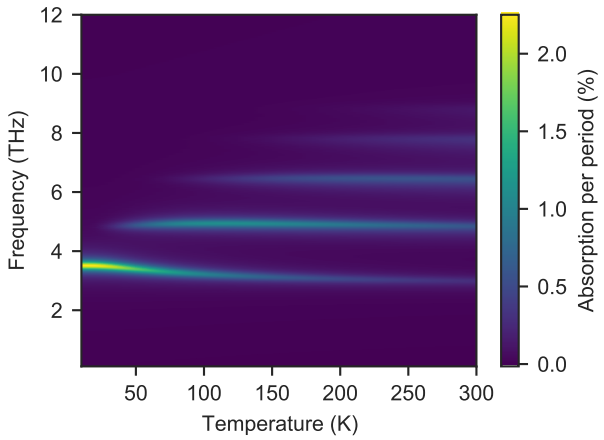

Figure 5.5: The image shows the spectrum of a square well as a function of temperature. The simulation is performed under the exact same conditions as Fig. 5.4, but we have now filled in the temperatures in between to see the evolution of the absorption lines.

the doped system, we see an absorption resonance very close to 3.0 THz.

What has happened here is, the effects of plasmon-plasmon coupling have almost exactly "cancelled out" the effects of band-bending due to charge redistribution. After all the calculations, we end up with a single MSP mode, which is effectively locked to the bare (undoped) transition energy.[3] This cancellation is closely related to a well-known effect in cyclotron systems called the Kohn theorem [107], which has also been extended to semiconductor systems [108]. Ideally, The effect is derived in a parabolic potential of infinite extent. For PQWs in semiconductors, where we are forced to truncate the potential at some point, the effect will essentially remain as long as the PQW is made deep enough to accommodate all the carriers. That is, the maximum barrier composition must be chosen high enough that most carriers will not be excited into the continuum above the well.

This picture already explains why PQW are a natural choice for the study of THz polaritons compared to square wells, and indeed they have been used already to exhibit strong coupling at room temperature [67]. However, there is a complication to the square well's side of the story which is worth mentioning. In Figs. 5.4 and 5.5 we have used only a moderate modulation doping at 1×10^{11} cm^{-2} per well. At this doping level, the effect of multisubband interactions is present, but it is not strong enough to cause significant qualitative changes in the absorption behaviour of the square well. On the other hand, MSP theory predicts that very heavily-doped square QWs can exhibit a robust absorption peak even at room temperature [109], and this has been successfully used for ultra-strong coupling studies in the mid-IR [74]. So why can we not exploit this same effect in the THz?

Fig. 5.6, which shows the QW absorption at room temperature with higher doping levels, gives us a clue. For the square well, we see that at a doping level of 3×10^{11} cm^{-2}, the absorption peak is still split into several weak peaks. If we increase the doping level to 3×10^{12} cm^{-2}, we see that the peaks coalesce

[3] One might note that the PQW absorption frequency does not quite appear "locked" in Fig. 5.4, as there is a redshift from 30 K to 300 K. however this is largely due to bandgap variation with temperature, and it would appear in the bare transition frequencies as well.

into a single, strong absorption peak (similar to Ref. [109]). However, crucially, the peak is blue-shifted to around 12 THz – well above our original target of 3 THz. Loosely: to achieve a strong MSP resonance, we must dope the well so heavily that its plasma frequencies $\omega_{P,\alpha}$ are too large, and depolarization-like effects push the resonance out of the THz.

Interestingly, we also see that the PQW behaviour starts to deviate – blue-shifting and inhomogeneously broadening – when the doping level is heavy enough. As we mentioned above, the robustness of the PQW peak only survives as long as the well is deep enough to contain all the charges.[4] There is a rough rule of thumb which guides this "crossover" in behaviour of both types of wells. We can define an overall plasma frequency for the QW as

$$\omega_P^2 = \frac{e^2 N_{2D}}{\epsilon_0 \epsilon_s m^* L_{QW}} \tag{5.40}$$

where L_{QW} is the width of the QW (i.e., the region where electrons are confined, not including the barriers). Roughly, when this ω_P starts to exceed the bare transition frequency ω_{12} of the well, we cross over into a "heavy-doping" region with different absorption behaviour. In the PQW, when the doping is heavy enough that $\omega_P \sim \omega_{12}$, the Hartree potential from the electrons becomes so strong that it almost completely flattens the conduction band, and we approach a 3D plasma slab with a uniform charge density [108]. So, in fact, ω_P effectively sets an upper bound on the modulation doping that can be put into a PQW: beyond this, the additional electrons will stay bound to the Si-delta rather than settling the well. This is why bulk doping was used for the calculations of Fig. 5.6.[5]

This raises another point which is worth bearing in mind. To achieve the strong MSP resonance in a square well, it is imperative to use bulk doping rather than modulation doping. With the high doping densities required for this effect, band-bending due to the Hartree potential can become quite extreme. If we attempt to modulation dope a square well at 3×10^{12} cm^{-2}, the potential created by the positive charges of the Si-delta is so extreme that it creates its own QW around the delta, and electrons will not populate the square well. (It is not hard to imagine that this will be the case by looking at Fig. 5.2, where there is already a significant distortion of the conduction band edge for a square QW modulation doped at 3×10^{11} cm^{-2}.) If, on the other hand, we use a uniform bulk doping throughout the well, the uniform positive charge density of the Si ions adds a parabolic to the well, which counteracts the potential created by the electrons, helping to keep the electrons inside the well.[6]

Heavy bulk doping comes with a serious downside, though, which is not captured in our model: it puts a large reservoir of scatterers right in the middle of the well. This will significantly increase the effect of ionized impurity scattering, which could lead to a noticeable broadening of the linewidth. While, unfortunately, this is difficult to predict quantitatively – especially in the context of MSP modes – we saw in Ch. 4 that the relative linewidth of transitions can be a critical parameter for device design. So, all things considered, we see that square wells are problematic for ISB polariton studies in the THz regime. PQWs, on the other hand, provide a clean, stable resonance over a wide range of temperatures and doping levels.

[4] In the modulation doped case, as the doping increases, the Hartree potential from the electrons will make the PQW shallower and shallower until it cannot hold any more electrons. (We can see this "shallowing" begin to happen in Fig. 5.2.)

[5] Although note that, in the case of a PQW, we do have some control over ω_P because we can truncate the well at a higher composition to increase L_{QW}, thus increasing the maximum doping "capacity" of the well. If we had an infinitely-high PQW, we would not be limited in the number of charges we could accommodate.

[6] The fact that a uniform doping density creates a parabolic potential follows from the 1D Poisson equation $\partial_z^2 V = eN_{3D}/\epsilon\epsilon_0$, where N_{3D} is the 3D doping density and V is the electric potential. If N_{3D} is constant, then this equation is readily integrated, and we see that the potential V has to be of the form of a parabola $A + Bz - (eN_{3D}/2\epsilon\epsilon_0)z^2$, with A, B being constants determined by the boundary conditions. When written in terms of the potential energy of an electron ($-e \cdot V$), this takes the form of a parabolic potential well with curvature related to the squared plasma frequency.

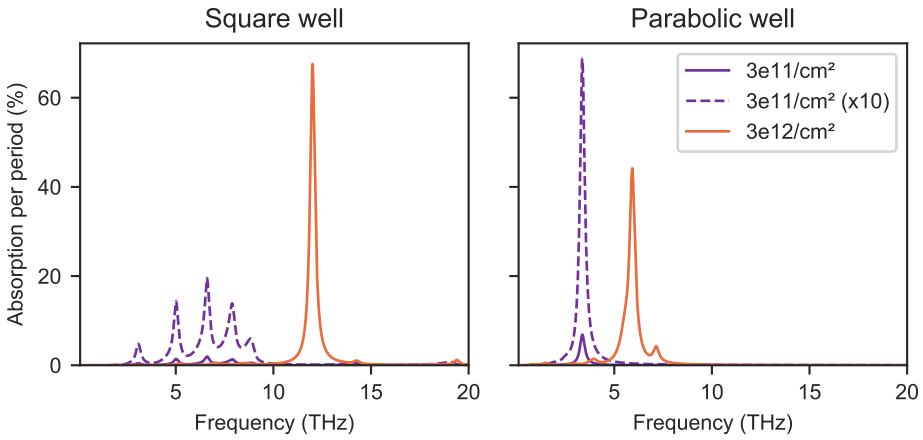

Figure 5.6: Absorption spectra (as defined in Sec. 5.5) for a square QW system (left) versus a PQW system (right), this time showing the dependence on doping at room temperature (300 K). Again, these are the same well systems described in Fig. 5.1, which are designed to absorb around 3 THz at low temperature. However, in this case we have *bulk* doped uniformly throughout the well at 3×10^{11} cm^{-2} and 3×10^{12} cm^{-2} per well. Since the absorption strength naturally increases with doping, the 3×10^{11} cm^{-2} is shown both at real scale (solid) and at 10× scale (dashed) for comparison. As before, we have used a constant $\gamma = 0.3$ THz for all calculations.

5.8 Some real-world considerations in active region design

So far in this chapter, we have focussed on the modelling and design on the scale of single QWs under periodic boundary conditions – effectively assuming an infinite stack. This is certainly the most challenging aspect of modelling an active region for THz polariton devices, however it is worth discussing some higher level design considerations as well.

We have already touched upon the risk of bulk-doping QWs due to ionized impurity scattering, but there are other challenges associated with doping. In Si modulation-doped $Al_x Ga_{1-x} As$ QWs, one must be careful of several other effects. One such effect is the segregation of Si during molecular beam epitaxy (MBE) growth, which can cause a significant asymmetric smearing of the doping delta [93], depending on the growth temperature. If the Si-delta is placed too close to the wells, Si atoms could penetrate into the well leading to an increase in ionized impurity scattering.

Another doping effect to keep in mind is the creation of deep-donor states when modulation doping into $Al_x Ga_{1-x} As$ barriers. When the $Al_x Ga_{1-x} As$ composition exceeds $x > 0.2$, the Si donor state binding energy can increase significantly due to the creation of DX-centres [110]. This can lead to poor ionization of the Si dopant atoms, leading to a lower-than-expected density of electrons in the QW. One mitigation strategy is to place a very small GaAs square well into the barrier so that the dopants can be placed in GaAs. Another strategy, employed in high-mobility two-dimensional electron gass (2DEGs) is to illuminate the sample prior to any measurements to break the electrons free from the metastable DX traps [110].

Finally, one last consideration when designing a structure is the surface and interface states which may be created at the surface of the sample and at the interface between the substrate and the epitaxial growth. If the structure is not designed carefully, these interface states can "steal" electrons from the QW region. This can lead to a depletion of the wells near the top and/or bottom of the QW stack, along with band-bending which could shift their transition frequencies compared to the wells near the centre of the stack. To mitigate this, ideally, the QWs should be placed far enough away from the surface and from the substrate interface. It may also be desirable to add additional doping at the edges of the QW stack to compensate for these surface and interface states. In principle, the exact compensation needed can be calculated by Schrödinger-Poisson of the entire structure, treating the surface and substrate interface as Schottky barriers. Unfortunately, it is difficult to predict the exact barrier height, as this can depend significantly on the details of the epitaxial growth and the fabrication. For certain structures, overcompensating can be problematic, and it may take experimentation to get the balance just right [111].

Chapter 6

Molecular beam epitaxy: some fundamentals

Molecular beam epitaxy (MBE) has frequently underpinned the development of quantum and nanotechnologies over the last few decades. As these technologies become more advanced, so too are their demands for high-precision crystal growth. For example, our recent achievement of a terahertz (THz) quantum cascade laser (QCL) with record-high operating at 250 K, required hundreds of $Al_x Ga_{1-x} As$ layers with a total growth time approaching 20 hr [15]. Although it remains unclear exactly to what extent the MBE structure quality plays a role in the performance of these quantum cascade devices, there is growing evidence to suggest that very high precision is necessary [112]. While MBE is far from the only factor in making these devices successful – structure design and post-MBE fabrication are both critical as well – prudent growers should take every available precaution to make sure that the structure they grow is as close as possible to the intended design.

This becomes even more challenging for the THz intersubband (ISB) polaritonic devices discussed in this thesis, which we have seen may require parabolic quantum wells (PQWs) of high quality factor. In the next chapter, we will look at novel MBE growth techniques for such PQWs. In this chapter, however, we will discuss a selection of considerations for growing high-quality $Al_x Ga_{1-x} As$ structures in general. Many of the results will be broadly applicable in MBE, although our focus will be on III-V semiconductors. While these strategies are not necessarily novel, they are not universally employed by the MBE community, and they include approximations which are not always well-examined. We will take the opportunity to clarify and justify some of these approximations. At the same time, these basic techniques of flux calibration will be an important foundation for our work on continuous alloys in the next chapter.

6.1 The relationship between growth rate and flux

The basic idea of MBE is to grow crystals by shooting particles (usually atoms or simple molecules) at the surface of a starter crystal, called a substrate. (See Fig. 6.1, for example.) Compared to other growth techniques, MBE uses relatively slow growth rates in the range of a few Å/s or less, and is performed in a very clean ultra-high vacuum (UHV) environment. The advantage – one of the hallmarks of MBE – is that we obtain extremely precise control over the composition and thickness of the crystal layers being grown.

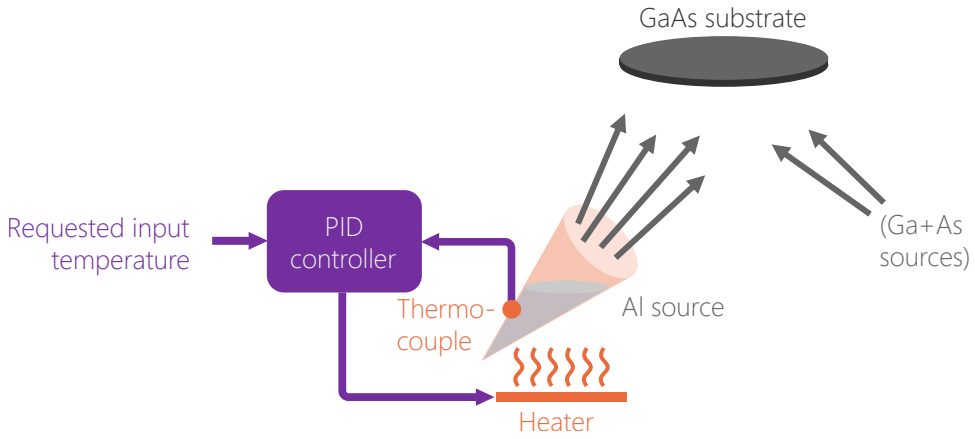

Figure 6.1: Conceptual diagram of an AlGaAs MBE growth. We see the Al crucible which contains molten material, with temperature controlled by a proportional, integral, derivative (PID) feedback loop.

This makes it one of the best techniques to grow nanometre-scale 1D heterostructures, particularly if quality and precision are the most important factors.

To achieve such precision in layer thicknesses and control over the composition, though, one of the key quantities that an MBE grower must pay attention to is particle flux. To grow a layer of a certain target thickness, the MBE grower needs to know how fast the layer is growing so that they may stop the growth at the appropriate time. (In practice, of course, this usually means programming a computer to stop the growth at the right time.) For a single-element material, the growth rate G has the following relationship with the particle flux:

$$G = \frac{\left(a_l^{\parallel}\right)^2 a_l^{\perp}}{Z} \cdot s \cdot \phi \tag{6.1}$$

Here ϕ is the flux of atoms impinging on the substrate per unit time, per unit area. s is the sticking coefficient: the fraction of those atoms that actually end up sticking to the substrate. Z is the number of atoms per unit cell. $a_l^{\perp}$ and $a_l^{\parallel}$ are the in-plane and growth-direction lattice constants of the layer being grown, and they must be calculated to account for the strain of the grown layer.[1]

Suppose that the grown material naturally (i.e., in the absence of stress) has a lattice constant a_l.[2] When this material is grown *epitaxially* on a substrate, the natural lattice constant of the grown material will stretch or shrink in-plane to match the substrate lattice constant, a_s. So we have $a_l^{\parallel} = a_s$. In the growth direction, the lattice constant will then tend to shrink or stretch to compensate. This can be calculated

[1] We can gain some intuition for Eq. (6.1) by grouping terms. $(s \cdot \phi)$ is the flux of atoms that stick to the substrate, and $\left(Z \Big/ \left(a^{\parallel}\right)^2 a^{\perp}\right)$ is the number of atoms per unit volume in the grown material. So the growth rate is simply equal to the number of atoms sticking to the substrate per unit time, per unit volume, divided by the number density of the grown material. Importantly, the number density is a fixed property of the material being grown, while the flux is controlled by the MBE grower. So we see that there is a very direct relationship between flux and growth rate.

[2] Strictly speaking, there just needs to be a periodic relationship between atoms of the substrate and atoms in the grown layer. As long as the atoms "match up," the crystals could have different lattice constants or could even be oriented in different directions. However, here we assume the simplest case of a 1:1 match-up of the atoms.

knowing the material's Poisson ratio, ν_l:

$$a_l^{\perp} = a_l + 2\frac{\nu_l}{1-\nu_l}(a_l - a_s) \tag{6.2}$$

In some cases, the grown layer will *relax*, meaning that its in-plane lattice constant will deviate from that of the substrate back towards its natural value. In that case, the calculation of lattice constants will need some modification, but Eq. (6.1) will still hold.

While this description has been limited to a single-element material, it is readily generalized to compounds with fluxes of several materials. One must just take care to calculate the number density and sticking coefficients correctly for the compound. For the III-V semiconductors grown in our lab (compounds of Al, Ga, In, As, Sb), it turns out that the growth is determined by the group III fluxes (Al, Ga, In) under most conditions. As long as *enough* group V (As or Sb) is supplied to grow the material, any excess will simply desorb off the substrate, ensuring layer stoichiometry. Thus, the growth rate is controlled by the sticking coefficient and fluxes of the group III elements alone. The group V fluxes may have some influence on the growth kinetics and final layer quality, but they do not need to be controlled very precisely, except in cases where more than one group V element is being used, in which case their ratio will be important for the final composition of the material. (We will not discuss those situations here, however.)

6.2 Producing atomic flux with an effusion cell

In MBE, the most common way to generate fluxes is by using *effusion cells*. Effusion cells have a crucible filled with a single pure element (or sometimes a pure compound), with an opening facing the substrate. When the material in the crucible is heated up, it begins to evaporate and release atoms through the opening. In the UHV environment of an MBE chamber, these evaporated atoms have a very long mean-free-path, so they travel unimpeded to the substrate – acting as a beam of particles.

While some cells have a valve to modulate the flux of particles escaping the cell, most group III cells only have a shutter with two states: fully open, or fully closed. In that case, the MBE grower controls the flux by varying the temperature of the crucible: higher temperature leads to higher flux. So the question is: what temperature should a grower set the cell to, to obtain a desired flux (and thus a desired growth rate and composition)?

Assuming an ideal gas at equilibrium, the particle flux, ϕ, can be calculated in relation to the vapour pressure, P, of the material in the crucible [113]:

$$\phi = \frac{P_v(T)}{\sqrt{2\pi m k_B T}} \tag{6.3}$$

Here m is the mass of the particles emitted, k_B is the Boltzmann constant, and T is the absolute temperature of the material. In reality, this is a non-equilibrium situation creating a non-isotropic beam of particles, but the beam flux hitting a given point on the substrate will at least be proportional to this value. Furthermore, the vapour pressure of metals tends to be given in the form of an Antoine relationship, $\log(P_v) \propto A - B/T$. These facts allow us to write down an empirical relation between the effusion cell temperature, T, and the

flux, ϕ, hitting the centre of the substrate[3]:

$$\boxed{\phi = \frac{\alpha}{\sqrt{T}} \exp\left(-\frac{\beta}{T}\right)} \tag{6.4}$$

Here α and β are unknown constants. So, in principle, if a grower can calibrate α and β accurately, then they can produce any desired particle flux by setting the appropriate cell temperature.

The reality is a bit more complicated than this, however, as the coefficients α and β will tend to change over time. This can be due to a redistribution of molten/solid material within the crucible – e.g., due to the buildup of droplets, liquid creep up the side of the crucible, or simply the fact that the material is slowly used up over time. It can also be related to the fact that current technology does not allow us to accurately measure the temperature of the emitting material directly. Instead, "cell temperature" usually refers to the temperature of the *outer* surface of the crucible, which can be measured by direct contact with a thermocouple. While a PID feedback loop can be used to keep this thermocouple temperature stable at a desired temperature, this does not necessarily keep the emitting surface at a stable temperature, which can lead to instabilities in the flux.

Much of the job of limiting these instabilities falls to effusion cell designers (and MBE system designers in general). For example, in the hot-lipped "SUMO" cells for Ga and In, one of the key design goals is to minimize the formation of droplets on the crucible nozzle, which are a major cause of flux instabilities for these materials. Well-designed modern effusion cells can – with good luck – produce a flux which is stable to within 1 % or better over a period of 24 hr. Consistency and reliability, however, remain a challenge as the limits are continuously pushed. Even with an extremely stable effusion cell, one cannot expect the flux-temperature relationship to remain stable over a period of days, weeks, or months. Thus, it is crucial for MBE growers to be able to measure a cell's flux directly for frequent recalibration.

6.3 Measuring flux with an ion gauge

Hot filament ion gauges, like the one seen in Fig. 6.2, are designed to measure pressure. They are most commonly used to measure the extremely low pressures (down to the 10^{-12}–10^{-11} Torr range) in a UHV environment. It turns out, though, that these pressure gauges can be used – without any modification to the gauge head – to measure small particle fluxes. This is because ion gauges do not really measure pressure directly, but rather measure the density of particles.

An ion gauge works by ionizing some fraction of the particles inside a fixed volume (the cylindrical grid seen in Fig. 6.2). These ionized particles are captured by a thin voltage-biased collector filament, and by releasing their electrons generate a current that can be measured by circuitry designed to measure small currents in the picoampere range. This collector current is proportional to the number of particles in the collector volume at any given time. Under normal operation, this particle density is readily converted into a pressure via the ideal gas law. However, if the gauge is inserted into a beam of particles, this particle density measurement can instead be converted into particle flux.

This is one of the primary flux measurement strategies for many MBE labs, including our own. When the gauge is operating in this mode, we refer to it as a beam flux-monitoring ion gauge (BFM) (another

[3]Note that some sources do not include the $\sqrt{T}$ in this expression, and that a closer look at the underlying physics of evaporation reveals that the flux-temperature relationship may be even more complicated [114]. However, as shown in Appendix B.2.2, the use of different formulae in this case is approximately equivalent to a redefinition of the calibration coefficients. For this reason, this formula usually works well enough in practice, even if the $\sqrt{T}$ is neglected.

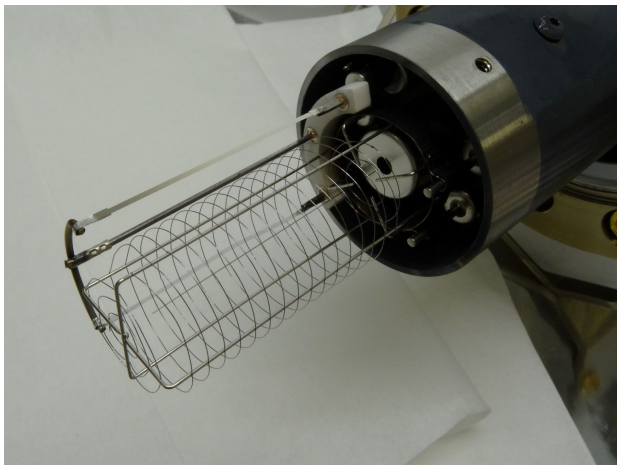

Figure 6.2: Image of a BFM gauge head from our Veeco GEN10 MBE system. This is a standard nude-filament ion gauge head for measuring UHV pressures. It is attached to an actuator so that it can be moved in front of the substrate holder as necessary.

common term is "monitoring ion gauge" or MIG). The BFM is placed on an actuator so that it can be moved in front of the substrate holder and probe the flux that would normally hit the centre of the substrate. This provides a fast, reliable flux measurement – the primary downside being that it cannot be performed during an actual growth because it blocks the flux to the wafer.

Further, some care must be taken when performing these measurements, as the relationship between particle density and particle flux depends on their speed. Imagine that a particle flux ϕ enters the gauge through a surface area A. If we define $n = \phi \cdot A$ as the number of particles entering the gauge per second, and P as the probability that a given atom will be ionized and collected, the collector current will be given by

$$I = nP \tag{6.5}$$

Let's assume that the ionization probability is proportional to the time, t, spent inside the gauge. Then $P = p \cdot t$, where p is the ionization probability per unit time. Further, let's assume that the time spent in the gauge is equal to $t = L/v$, where L is the average distance the particle travels through the gauge, and v is the average particle velocity. Putting this all together, the collector current is given by

$$I = \phi \cdot \frac{pAL}{v} \tag{6.6}$$

The quantities A and L are constant, since they depend on the geometry of the gauge. p will depend on the type of atom being measured, as well as the voltage biasing of the gauge, but we can assume that it does not change significantly over time. However, the velocity v will depend on the source of the particle beam we are measuring. If we assume that the beam is emitted by a hot surface – such as in an effusion cell – then the velocity will be proportional to $\sqrt{T}$, where T is the temperature of the emitting surface. Putting all this together, we find that the collector current is proportional to $\phi/\sqrt{T}$. Thus, we can introduce a

calibration constant, γ, and write:

$$\boxed{I = \frac{\phi}{\gamma\sqrt{T}}} \tag{6.7}$$

This γ factor is different for each effusion cell in an MBE system. It can be calibrated by taking a BFM reading of cell flux, and then immediately using that cell to grow a test layer. By measuring the actual thickness of the test layer – say, by optical reflectometry or high-resolution x-ray diffractometry (HR-XRD) – one can work backwards to determine what the actual flux was, corresponding to the BFM reading. Once γ is calibrated in this manner, the BFM provides a much more quick and convenient way to measure cell fluxes.

For practical purposes, it is often convenient to combine (6.7) with (6.4) to obtain the BFM reading as a function of cell temperature

$$I = \frac{\alpha}{\gamma T}\exp\left(-\frac{\beta}{T}\right) \tag{6.8}$$

From this we obtain a convenient way to calibrate the α and β coefficients if we know γ. We perform a series of BFM measurements $I[i]$ at different temperatures $T[i]$. Plotting $\ln(I[i]T[i])$ versus $-1/T[i]$ should yield a straight line with slope β and intercept $\ln(\alpha/\gamma)$.[4]

6.4 Analytical inverse of the flux equations

Quite commonly, an MBE grower may want to invert the flux-temperature relationship (Eq. (6.4)) or the BFM-temperature relationship (Eq. (6.8)) to find the cell temperature required to achieve a certain flux or BFM reading. However, inverting such equations is not within the standard mathematical toolbox of most MBE growers. Sometimes this leads growers to neglect the factors $1/\sqrt{T}$ and $1/T$ entirely. This approximation is not as drastic as it might seem at first glance – in Eqs. (6.4) and (6.8), dropping these factors is equivalent to a redefinition of the coefficients α, β, and γ. See Appendix B.2.2 for a more detailed explanation.

However, this approximation is somewhat insidious because it may lead us to implicitly drop the $\sqrt{T}$ factor from the BFM-flux relationship, Eq. (6.7). This particular $\sqrt{T}$ *cannot* be safely neglected in typical MBE growths, unless T is very close to the temperature, T_0, used for calibration. Neglecting this factor means that the growth rate calculated by a BFM measurement will be off by a factor of $\sqrt{T/T_0}$, which can easily be an appreciable error. If $T_0 \sim 10^3$ K and $|T - T_0| \sim 10^2$ K, we obtain a growth rate error around 5 %.

Fortunately, it is possible to invert equations like (6.4) and (6.8) without neglecting any factors by using the Lambert W function. While the Lambert W function may not be commonly known, implementations of it are readily available in most computational programming languages such as Matlab or Python. The Lambert W function is defined implicitly by the relation

$$W(ze^z) = z \tag{6.9}$$

[4]Note that this calibration procedure gives a value of α/γ rather than α directly. For this reason, the QNC-MBE lab uses a slightly different definition of the calibration coefficients in practice: see Appendix B.1 for details.

for real or complex z. This is useful because both Eqs. (6.4) and (6.8) have the general form

$$y = aT^{-r}\exp\left(-\frac{b}{T}\right) \tag{6.10}$$

which we can rewrite as

$$-\frac{b}{rT}\exp\left(-\frac{b}{rT}\right) = -\frac{b}{r}\left(\frac{y}{a}\right)^{1/r} \tag{6.11}$$

Then, applying $W(\cdot)$ to both sides, we can solve for T using

$$\frac{1}{T} = -\frac{r}{b}\cdot W\left(-\frac{b}{r}\cdot\left(\frac{y}{a}\right)^{1/r}\right). \tag{6.12}$$

We have to be careful here, though, because the Lambert function is multi-valued over certain ranges of arguments, and we need to use the correct branch. In the cases of interest to us, we have $T, b, r > 0$, so it turns out that the argument of the Lambert W function lies in the range

$$-\frac{1}{e} < -\frac{b}{rT}e^{-b/rT} = -\frac{b}{r}\left(\frac{y}{a}\right)^{1/r} < 0 \tag{6.13}$$

which is exactly the region where the Lambert function is multi-valued. In other words, there are actually two values of T which give the same value of y: one with $0 < T < b/r$ and another with $T > b/r$. In practice, we typically have b on the order of 10^4 K while T is much smaller, on the order of 10^3 K. So, we want the solution with $0 < T < b/r$, which corresponds to the *lower* branch of the W function, denoted as $W_{-1}(\cdot)$.

So, given $y = aT^{-r}e^{-b/T}$, we have

$$\frac{1}{T} = -\frac{r}{b}\cdot W_{-1}\left(-\frac{b}{r}\cdot\left(\frac{y}{a}\right)^{1/r}\right); \quad r > 0 \tag{6.14}$$

It must be emphasized that this branch choice only applies to exponents $r > 0$. If $r < 0$, then the Lambert function is single-valued and the lower branch does not exist. If $r = 0$, then the equation reduces to the simpler form $y = ae^{-b/T}$, which has solution

$$\frac{1}{T} = \frac{1}{b}\ln\left(\frac{a}{y}\right); \quad r = 0. \tag{6.15}$$

6.5 A typical flux calibration procedure

By now, we have the key equations necessary for a typical III-V MBE growth. To illustrate how everything fits together, it is useful to walk through a typical pre-growth flux calibration procedure, where we assume that the γ coefficient is well-calibrated but the α and β coefficients may have drifted. Drift of these coefficients is quite common, e.g., when cells are reduced to an idling temperature overnight and then brought back up for a growth. (Again, for readers from the QNC-MBE lab, note that the α, β, γ coefficients used here are slightly different from those used in the lab. See Appendix B.1.)

Suppose we want to grow a 1000 Å layer of AlAs on a GaAs substrate. We know that the AlAs growth rate is determined by the Al flux, as long as we supply sufficient As, so our first step is to figure out what the

Al flux should be. In more complicated growths with several types of material, there may be constraints, but here we are essentially free to choose the Al flux. We might decide we want to grow at $G = 1\,\text{Å/s}$ for 1000 s to achieve the desired thickness. Then, based on this growth rate, we can calculate the flux using Eq. (6.1). For AlAs at typical growth temperatures, we can assume that the sticking coefficient is 1 and that the number of Al atoms per unit cell is $Z = 4$, so

$$\phi = \frac{4G}{a^{\perp}(a^{\parallel})^2} \tag{6.16}$$

which works out to $\phi = 5.519 \times 10^{13}\,\text{cm}^{-2}\text{s}^{-1}$.

The next question is, what temperature should we set the Al cell to to achieve this flux? Here we use Eq. (6.14), with $y = \phi$, $a = \alpha$, $b = \beta$, and $r = 1/2$:

$$T = -2\beta\left[W_{-1}\left(-2\beta \cdot \left(\frac{\phi}{\alpha}\right)^2\right)\right] \tag{6.17}$$

If we use some typical values, e.g., $\alpha = 1.5 \times 10^{28}\,\text{K}^{1/2}\text{cm}^{-2}\text{s}^{-1}$ and $\beta = 40\,000\,\text{K}$, we obtain a cell temperature of 1349.9 K or 1076.8 °C.

However, it is important to note that this is only the correct temperature if the α, β coefficients have not changed since the last calibration. In practise, we should treat this as a first "guess" for the cell temperature. Then, once the cell is at this temperature, we should measure the flux using the BFM – its reading will be converted into flux using Eq. 6.7. We get

$$\phi = \gamma I \sqrt{T} \tag{6.18}$$

(Again, assuming that γ has already been well-calibrated.) If the flux isn't at its desired value, we adjust the cell temperature (there is a useful approximation for doing this, outlined in Appendix B.2.1), and then keep repeating the flux measurements until we reach the value we want.

It is important to note that with this procedure, the α and β coefficients end up having *no* impact on the final accuracy of the growth. They are only used to get us an initial guess for the temperature, which is then corrected iteratively using the BFM. In the end, the accuracy of the growth is completely dependent on the accuracy of the BFM and, in particular, its calibration coefficient γ. Of course, the picture may get murkier in some cases – for example, in the continuous alloys discussed later. Furthermore, it should be noted that an ion gauge is not the only way to measure flux, so this is certainly not the only procedure one could follow. Nevertheless, this emphasizes the point that accurate flux measurement is a critical component of high-quality MBE growth.

6.6 Difficulties in defining cell temperature

In the discussions above, we have been somewhat imprecise with the definition of temperature, T. Strictly speaking, for these derivations to hold, T should be the temperature of the emitting surface. And further, there might be several emitting surfaces within the same cell – each potentially at a different temperature. Still, even if we imagine that all the flux comes from the surface of the molten metal in the crucible, the temperature of this surface cannot be directly measured in practice. We are only able to reliably measure the temperature of thermocouples placed outside the crucible, and we know that these thermocouple temperatures could be quite different from that of the emitting surface. Effusion cells typically operate

under large thermal gradients, such that thermocouples placed at the tip and base of the crucible will often have readings that differ by more than 100 °C, with the molten emitting surface likely lying at some temperature in between. It is not difficult to imagine, then, that the emission temperature relevant for flux calculations could differ from the measured thermocouple temperatures by 50 °C or more.

At first glance, this appears disastrous, because cell fluxes are quite sensitive to temperature – depending on the element, a change of just 1 °C can lead to several percent change in the flux. So, how are we able to obtain a reliable flux when the uncertainty in the emitting surface temperature is so large? As it turns out, the calibration procedure for the γ coefficient is the key here. This calibration is done in such a way that it implicitly eliminates much of the error due to temperature readings – so long as that error is constant over time.

Suppose our measured temperature, T_m, is offset from the actual material temperature T_a by some constant offset ΔT.

$$T_m = T_a + \Delta T \tag{6.19}$$

Now suppose we perform a calibration procedure to measure γ: we set the cell temperature to some temperature $T_{m,0}$, and we measure the corresponding flux, ϕ_0, and BFM reading, I_0. If we use $T_{m,0}$ as the temperature in Eq. 6.7, then we will calculate γ to be

$$\gamma_m = \frac{\phi_0}{I_0\sqrt{T_{m,0}}} \tag{6.20}$$

but the actual γ should be

$$\gamma_a = \frac{\phi_0}{I_0\sqrt{T_{a,0}}} \tag{6.21}$$

where $T_{a,0}$ is the actual temperature during calibration.

Going forward after the calibration, we continue to use the measured temperature, since we do not know the actual temperature, so we will convert BFM readings into flux values using the formula

$$\phi_m = I\gamma_m\sqrt{T_m} \tag{6.22}$$

when the actual flux is

$$\phi_a = I\gamma_a\sqrt{T_a} \tag{6.23}$$

With some algebra, it can be shown that the relative difference between these two quantities is

$$\frac{\phi_m - \phi_a}{\phi_a} = \sqrt{\frac{(1 + \Delta T/T_a)}{\left(1 - \Delta T/T_{a,0}\right)}} - 1 \tag{6.24}$$

Using the binomial approximation, this becomes

$$\frac{\phi_m - \phi_a}{\phi_a} \approx \frac{\Delta T}{2}\left(\frac{1}{T_a} - \frac{1}{T_{a,0}}\right) \tag{6.25}$$

which can be further simplified to

$$\frac{\phi_m - \phi_a}{\phi_a} \approx \frac{\Delta T \cdot (T_{a,0} - T_a)}{2T_{a,0}^2} \tag{6.26}$$

So this error is second-order in the temperature differences: ΔT (thermocouple offset) and $T_a - T_{a,0}$ (difference between melt temperature during current growth vs melt temperature during calibration growth.)

Typically, we'd have $T_a, T_{a,0}$ on the order of 10^3 K. In a worst-case scenario, the thermocouple offset might be $\Delta T \sim 10^2$ K, and the cell temperature T_a used for a particular growth might be 10^2 K different from the temperature used for calibration. Even in this extreme case, the relative flux error due to this temperature offset is only 0.5 %. If the cell temperature is close to the temperature used for calibration, the error will be even smaller. So we see that the error turns out to be quite small. Essentially, the calibration procedure implicitly redefines γ so that the equations give the correct flux, as long as the cell temperature does not deviate too far from the calibration point.

Of course, this is only an approximate picture, and we have neglected the possibility that a cell could include several emitting surfaces at different temperatures. Still, the same argument should apply, so long as the relative contributions of those surfaces remain constant over time. The idea is that γ is automatically selected to give the most accurate flux reading in spite of all the temperature uncertainties.

That said, this is still not the full story, as we can expect that the cell will change over time, leading to some stochastic variation in ΔT from day-to-day – call this δT. This will cause the flux measured by BFM to be off by

$$\frac{\phi_m - \phi_a}{\phi_a} \approx \frac{\delta T}{2T_a} \tag{6.27}$$

This is a first-order error term, but the important thing is that we expect the random variations δT to be much smaller than the constant offset ΔT. If δT is around 2 K and T_a around 1000 K, we obtain an error of just 0.1 %. On the scale of a single growth, we expect δT to be much smaller than this, since otherwise the flux would be unstable anyway. However, these errors may still be worth considering in the long term. If ΔT drifts significantly over a long period of time (e.g., as the material in the crucible is used up), it may be necessary to perform a recalibration of the γ coefficient in order for the BFM to be a reliable flux gauge. (And, of course, there may be other motivations to recalibrate γ anyway, such as a change in the beam geometry over time.)

Finally, note that a similar principle applies to the α and β coefficients, as discussed in Appendix B.2.3. Our imperfect knowledge of the actual temperature of the emitting surface largely amounts to an implicit redefinition of the α, β coefficients. Usually, this is not as critical as the γ coefficient, since the α and β coefficients are only used to set an initial guess for the cell temperature. However, there are certain cases where it is impossible to measure the flux directly before growth (e.g., if a cell temperature must be changed mid-growth). In these cases, it will be more important to ensure that α, β and their related equation are treated correctly.

6.7 Conclusions

In this chapter, we have discussed flux calibration procedures in MBE. Most important for our efforts in the next chapter are the equations Eqs. (6.4) and (6.7), which give us the relationships between flux, BFM

reading, and cell temperature. We have also discussed how to mathematically invert these equations, and outlined a typical pre-growth flux measurement procedure. Finally, we have discussed some issues related to the approximations often used by MBE growers and the difficulties in defining cell temperature correctly. These issues can be somewhat difficult to keep track of – we see that there are several cases in which calibration coefficients may be implicitly redefined to mitigate an underlying approximation or uncertainty. In many cases, the end result will work out just fine, but this is not universally true. It is therefore important for MBE growers in high-precision applications to understand when approximations are being implicitly applied and when they can be justified.

Chapter 7

Graded alloy growth with molecular beam epitaxy

Note: significant portions of this chapter are derived from our publications, Refs. [115] and [116], with permission from the publishers.

7.1 Introduction and background

Molecular beam epitaxy (MBE) has long focused on the growth of discrete layered systems with sharp interfaces. The concept of stacking different materials is a simple but versatile concept, enabling a dizzying array of physics and technology. Even when just two or three materials are used, enormous complexity can be achieved by employing a large number of layers with carefully engineered thicknesses. Striking examples have existed for decades now, such as quantum cascade lasers (important emitters in the mid-infrared (IR) [117] and terahertz (THz) [22]), which routinely incorporate hundreds of layers into a single device.

Still, while discrete layers have dominated the landscape of 2D heterostructures, they represent only a small fraction of the possible design space. Graded structures with smooth variations in alloy composition have been much less explored. If arbitrary composition profiles could be generated along the growth direction, one could easily envision growing exotic quantum well and barrier structures with unique, unexplored physics. However, such structures have remained difficult to grow, since they require precise time-varying fluxes that most MBE systems are not designed to generate. If time-dependent fluxes could be controlled more precisely – and, especially if this could be done without major modifications to the MBE system – it would open up new realms of structure design.[1]

Furthermore, developing such techniques to grow graded alloys could also be useful in preventing undesirable grading during regular growths. MBE growers often encounter situations where the composition is unintentionally and undesirably graded – for example, because of thermal transients when opening and closing the cell shutters. While modern MBE systems are designed to minimize these transient effects to

[1] Strictly speaking, graded alloys can be grown with other techniques like metal-organic vapour-phase epitaxy (MOCVD) or chemical beam epitaxy (CBE), but these techniques do not offer as high precision as MBE, and have increased background doping levels which can be detrimental when high-quality structures are needed.

the point that they are negligible for most growths, these effects may become more and more important as the demands for device performance and reproducibility continue to grow. THz quantum cascade lasers, again, are a prime example of this – the race towards room temperature operation (recently capped [15] at 250 K) is underpinned by complex growths which demand the utmost precision. Thin quantum well barriers are particularly susceptible to shutter transients, and when the growth introduces uncertainties like this, it only serves to complicate a design and modelling problem that was already challenging to begin with. Techniques to gain better control over cell fluxes may be of great importance in such demanding situations.

In this chapter, we explore time-dependent flux control in MBE, focusing on the example of $Al_x Ga_{1-x} As$. $Al_x Ga_{1-x} As$ is well-studied and particularly convenient for graded alloys, since it is well lattice-matched to GaAs substrates at all composition values, and it has a linear bandgap dependence [118] for $x < 0.38$. Thus, by grading the composition profile, x, it is relatively straightforward to engineer the exact potential energy profile of a quantum well or barrier.

We study the relatively simple example of quadratically varying parabolic quantum wells (PQWs), which can already prove useful in practice. PQWs have equally spaced subbands and an optical response which is strikingly different from the more ubiquitous square quantum well. The optical response of PQWs is robust to both charging effects [108] and the redistribution of electrons at higher temperatures [109] – characteristics which have recently made them attractive in the study of strong light-matter coupling in the THz [60], where square wells are limited to cryogenic temperatures [72]. PQWs could thus be a key ingredient in the development of practical polaritonic devices within the so-called "THz gap [59]." However, to make these devices a reality, growing graded quantum wells of the highest quality could prove essential.

The idea to grow PQWs in $Al_x Ga_{1-x} As$ is not a new one, nor is the idea to grow continuously graded quantum wells in general. However, applications have been limited as it remains difficult to achieve the necessary time-varying control over the atomic flux. The easiest and most common way around this is to use so-called "digital alloys" – rapidly actuating the cell shutter to provide short pulses of flux. By varying the duty cycle, one can approximate a continuous alloy. Indeed, PQW studies to date (both early on [119] and more recently [60, 88]) have largely employed such digital alloys. Unfortunately, since shutters typically cannot be actuated more than a few times per second, the quality of digital alloys is inherently limited: they cannot exactly capture the target composition profile, and they generate a large number of interfaces. Further, while faster shutter action should, in principle, lead to a better approximation of the continuous alloy, it comes with drawbacks such as flux transients that need to be carefully compensated [120]. Rapid shutter actions (and thus, short pulse times) also make it difficult to achieve good in-plane uniformity unless the shutter actions can be precisely timed to the substrate rotation.

So, for the highest-quality structures, it would appear most promising to pursue a smooth composition profile. This requires accurate time-dependent control over the cell flux, which is difficult due to the complicated thermal behaviour of effusion cells [121]. Attempting to develop an accurate model of an effusion cell's thermal dynamics is highly impractical, and depends on many system-specific factors. Many MBE systems mitigate these dynamical effects via "black box" proportional, integral, derivative (PID) temperature controllers, but these controllers are optimized only for the steady-state case. In the dynamic case required for graded structures, these PID controllers are only sufficient when the structure can be grown slowly enough that the effusion cell remains in a quasi-equilibrium [122, 123]. Unfortunately, such slow growth rates lead to long growth times, which make the growths more susceptible to random flux deviations, and limit the total thickness of the structures that can be grown. This means that PID control is not sufficient when applications demand graded quantum well superlattices of several micrometres, for example.

A natural inclination might be to design a more sophisticated feedback controller to better handle the dynamic case. Improving the responsivity of the temperature controller can indeed yield some benefit [124], but there are limits on accuracy. Unfortunately, while the typical crucible temperature measurement is a good proxy for atomic flux in the steady-state, it does not measure the surface of the molten material directly, so it may lead or lag the flux in the dynamic case. Thus, the most challenging part of implementing a dynamic feedback controller lies in monitoring the atomic fluxes accurately in real time during a growth – something that most MBE systems are not equipped to do.

In the direction of feedback control methods, perhaps the most promising work was by Aspnes *et al.* [125], using spectroscopic ellipsometry for real-time composition measurement. Still, such ellipsometry systems remain uncommon in MBE labs, and they are difficult to set up and calibrate accurately. A lack of substrate rotation further means that the growth will be extremely non-uniform across the surface of the substrate.

Another option is to "flip" the problem – keeping the cell fluxes constant and varying the desorption rate of Ga off the substrate. The desorption can be monitored with a mass spectrometer, and it can be controlled via the substrate temperature or the As overpressure, giving real-time control over the composition [126, 127]. This method requires very high substrate temperatures, however, and is extremely sensitive to temperature non-uniformity across the substrate, again leading to poor lateral uniformity of the growth.

In the absence of real-time feedback information, one must try to find the appropriate cell inputs ahead of time. Several techniques in this vein have been attempted with varying degrees of success. An iterative guessing approach can be simple to implement, but requires several trials and may not track the desired profile very accurately [128]. More sophisticated approaches have used feedback loops under "test" conditions to determine the appropriate input to the Al cell [120, 129]. This takes advantage of the fact that a beam flux-monitoring ion gauge can be placed in front of the substrate to accurately measure time-varying atomic fluxes coming out of an effusion cell – the main catch being that it cannot be done during a real growth. Such approaches are promising, but have still struggled to achieve highly precise tracking of smooth composition profiles at the growth rates required for thick structures.

Our initial approach [115] was in a similar direction – we used a flux-monitoring ion gauge to empirically derive a simple linear model of the Al cell's thermal dynamics. This model could be used to predict the required cell input for arbitrary composition profiles. In the approach detailed below, we improve this technique significantly, simplifying the linear model to make it more robust against measurement noise and adding compensation for shutter transients. Most importantly, we add an iterative correction approach which can reduce errors in the linear model without the need for more complex modelling. This can be used on growths that are most demanding on compositional control.

7.2 Converting composition to flux

We begin with the general problem of growing $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$ with smoothly varying composition in the growth direction, z. In MBE, we cannot control the material composition directly – it is the atomic fluxes that we must vary as a function of time to achieve the desired composition. So, our first step is to convert the target composition profile, $x(z)$, into Al and Ga fluxes $\Phi_{\mathrm{Al}}(t), \Phi_{\mathrm{Ga}}(t)$ as functions of time, t. Assuming 100 % sticking coefficients for Ga and Al, flux and composition are related by

$$x = \frac{\Phi_{\mathrm{Al}}}{\Phi_{\mathrm{Al}} + \Phi_{\mathrm{Ga}}}. \tag{7.1}$$

To keep things simple, we will choose a constant Ga flux, Φ_{Ga} and only vary the Al flux to control the composition. In that case, we have a target flux

$$\Phi_{\mathrm{Al}}(z) = \frac{x(z)}{1-x(z)}\Phi_{\mathrm{Ga}} \tag{7.2}$$

Note that this expression diverges when high Al content is required ($x \to 1$). This is not a problem for typical quantum well (QW) growths – including those studied here – because we work within the regime $x \leq 0.4$ where $\mathrm{Al}_x\,\mathrm{Ga}_{1-x}\,\mathrm{As}$ is direct-bandgap. If a wide range of compositions were needed within the same growth, the techniques discussed in this chapter could still be applied, but one would need to vary both the Ga and Al flux at the same time. (In that case, a sensible constraint might be to keep the total growth rate constant.)

The above expression for the Al flux is given as a function of position, but for a growth we need flux as a function of time. To convert position into time, we use the growth rate,

$$G_{\mathrm{AlGaAs}}(z) = \frac{\mathrm{d}z}{\mathrm{d}t} \tag{7.3}$$

Integrating, we get a relationship

$$t(z) = \int_0^z \frac{\mathrm{d}z'}{G_{\mathrm{AlGaAs}}(z')} \tag{7.4}$$

which can be numerically inverted to find $z(t)$. The growth rate depends on composition, and thus on z, so we have

$$G_{\mathrm{AlGaAs}}(z) = \frac{a_\perp(z)a_\parallel^2}{4}[\Phi_{\mathrm{Al}}(z) + \Phi_{\mathrm{Ga}}] \tag{7.5}$$

$$G_{\mathrm{AlGaAs}}(z) = \frac{a_\perp(z)a_\parallel^2}{4}\frac{\Phi_{\mathrm{Ga}}}{1-x(z)} \tag{7.6}$$

where $a_\perp$ is the growth-direction lattice constant of $\mathrm{Al}_x\,\mathrm{Ga}_{1-x}\,\mathrm{As}$ on GaAs, and $a_\parallel$ is the substrate lattice constant. So in this case we have $a_\parallel = a_{\mathrm{GaAs}}$ and, using Vegard's law,[2]

$$a_\perp(z) = x(z)a_{\perp,\mathrm{AlAs}} + (1-x(z))a_{\mathrm{GaAs}} \tag{7.7}$$

where $a_{\perp,\mathrm{AlAs}}$ is the growth-direction lattice constant of pure AlAs grown on GaAs. Inserting this above, we get

$$G_{\mathrm{AlGaAs}}(z) = \frac{a_{\mathrm{GaAs}}^2}{4}\left(a_{\perp,\mathrm{AlAs}}\Phi_{\mathrm{Al}}(z) + a_{\mathrm{GaAs}}\Phi_{\mathrm{Ga}}\right). \tag{7.8}$$

With equations (7.2), (7.4), and (7.8), we can convert any target composition profile $x(z)$ into a target Al flux $\Phi_{\mathrm{Al}}(t)$, given a constant Ga flux Φ_{Ga}.

[2] Strictly speaking, Vegard's law does not apply to $\mathrm{Al}_x\,\mathrm{Ga}_{1-x}\,\mathrm{As}$, and a bowing parameter should be used [118]. The equations can be modified to include this, but they become slightly more cumbersome and the difference is negligible for our purposes.

7.3 MBE growth with linear dynamical modelling

So the problem is reduced to that of accurately generating a time-varying Al flux, $\Phi_{\mathrm{Al}}(t)$, which we will now consider. The exact details of how to do this may depend on the particular MBE system, but the general approach should be broadly applicable – even to other materials and effusion cell designs. The system used in this work is a Veeco GEN10 with Al evaporated from a 60 cc conical pyrolytic boron nitride (PBN) crucible. This Al cell has two independent heaters, but only the base heater is used during normal operation to keep the crucible lip cooler than the base and minimize Al creep. (This happens because Al, unlike Ga and In, wets the PBN surface.) When an input temperature is specified by the operator, a PID loop adjusts the heater's power supply voltage to maintain the base thermocouple at this temperature. (See Fig. 6.1 for a schematic representation.)

For typical growths using constant flux, we would generate a desired Al flux via the exponential relationship, which we discussed in the previous chapter

$$\Phi_{\mathrm{Al}} = \frac{\alpha}{\sqrt{T_{\mathrm{in}}}} \exp\left(-\frac{\beta}{T_{\mathrm{in}}}\right), \tag{7.9}$$

where T_{in} is the input temperature to the Al cell, and α and β are coefficients which are routinely measured via calibration growths[3]. This relationship only holds for constant-flux growths, though, where the Al cell is in steady-state. If a time-varying input, $T_{\mathrm{in}}(t)$, is applied, the thermal dynamics of the cell will lead to flux transients, which have a settling time on the scale of a few minutes or more. Thus, to accurately vary the flux as a function of time, we must compensate for the thermal dynamics of the effusion cell.

For simplicity, we leave the existing PID control system in place without any special adjustments, and try to find the "correct" sequence of input temperatures, $T_{\mathrm{in}}(t)$, which will give the desired flux, $\Phi_{\mathrm{Al}}(t)$. For this, we develop an empirical model of the cell dynamics.

To begin, we already know that there will be a significant non-linearity in the relationship $T_{\mathrm{in}}(t) \rightarrow \Phi_{\mathrm{Al}}(t)$, analogous to the static relationship Eq. (7.9). So, we take this static relationship as a starting point, using it to convert the instantaneous flux into an effective cell temperature, $T_{\mathrm{eff}}(t)$, defined implicitly as

$$\Phi_{\mathrm{Al}}(t) = \frac{\alpha}{\sqrt{T_{\mathrm{eff}}(t)}} \exp\left(-\frac{\beta}{T_{\mathrm{eff}}(t)}\right). \tag{7.10}$$

Or, inverting this expression, we can define the effective cell temperature explicitly as

$$T_{\mathrm{eff}}(t) = -2\beta \cdot \left[W_{-1}\left(-\frac{2\beta}{\alpha^2}\Phi_{\mathrm{Al}}^2(t)\right)\right]^{-1}, \tag{7.11}$$

where $W_{-1}(z)$ is the lower branch of the Lambert W function,[4] and α and β are the same coefficients used during steady-state growths. In other words, when the cell is producing a flux $\Phi_{\mathrm{Al}}(t)$ at a given time, $T_{\mathrm{eff}}(t)$ is the temperature that would be required to produce that flux *if* the cell was in steady-state. As a

[3] As discussed in the last chapter, strictly speaking, we should be using the surface temperature of the emitting material here rather than T_{in}. However, since that temperature is not known exactly, we rely on the external crucible temperature measured by the cell's thermocouple. Fortunately, this approximation – routinely employed in MBE – is not as drastic as one might expect at first glance. This is because the same approximation is also used when calibrating the α and β coefficients in the first place. As a result, α and β are automatically selected so as to minimize the impact of the approximation – as long as the temperature does not deviate too far from the temperature(s) used for calibration.

[4] The lower branch of the Lambert W function is defined by $z = W_{-1}(ze^z)$ with the constraint $z \leq -1$. As showed in the previous chapter, choosing the lower branch is equivalent to assuming that $T_{\mathrm{eff}} < \beta$.

result, when the cell is in steady state, $T_{\text{eff}}(t)$ should be equivalent to both the input temperature, $T_{\text{in}}(t)$, and the temperature measured by the cell's thermocouple. In the dynamic case, however, these three temperatures may deviate from each other substantially.

Another interpretation of this effective temperature is that we have simply converted the cell's flux, $\Phi_{\text{Al}}(t)$ into units of temperature. The purpose is to simplify the mathematical relationship by eliminating a known source of non-linearity: Eq. (7.9). We expect that the relationship $T_{\text{in}}(t) \rightarrow T_{\text{eff}}(t)$ will be simpler (i.e., less non-linear) than the original relationship $T_{\text{in}}(t) \rightarrow \Phi_{\text{Al}}(t)$ that we are trying to model.

So, the remaining task is to find a dynamical relationship between the input temperature, $T_{\text{in}}(t)$, and the effective temperature, $T_{\text{eff}}(t)$. In general, these quantities will be governed by non-linear heat transfer equations [121], but we approximate the relationship as linear, so that all the dynamical information can be described by a frequency-domain transfer function $H(f)$. Once this transfer function is known, the relationship between the input and effective cell temperatures is simply

$$T_{\text{eff}}(f) = H(f)T_{\text{in}}(f), \tag{7.12}$$

where $T_{\text{in}}(f)$ and $T_{\text{eff}}(f)$ are the Fourier transforms of $T_{\text{in}}(t)$ and $T_{\text{eff}}(t)$, respectively. This frequency-domain relationship is readily inverted. So, if we can find the transfer function of the Al cell, we can find the input temperature required to generate any desired flux profile.

7.4 Transfer function of the Al cell

In principle, the transfer function is straightforward to measure. We apply a dynamic input, $T_{\text{in}}(t)$, measure the response, $T_{\text{eff}}(t)$, convert to the frequency domain, and then solve for $H(f)$ in Eq. (7.12). In practice, though, we must take some care in how this is done. For one thing, real-world measurements will always include error and noise. For another, this system is not exactly linear, and we are only seeking to approximate it as linear under typical operating conditions.

So, as a test input, we apply sudden steps to the input temperature, within the range of typical cell temperatures used for growth: around 1250–1350 K in our case. Step inputs like this are simple to apply, and will excite the relevant dynamics so that they can be characterized. To measure the response, $T_{\text{eff}}(t)$, we use a beam-flux monitoring ion gauge, placed near the centre of the substrate manipulator. The flux can be calculated from the ion gauge current, $j_{\text{IG}}(t)$, using the relationship

$$\Phi_{\text{Al}}(t) = \gamma \sqrt{T_{\text{eff}}(t)} \cdot j_{\text{IG}}(t). \tag{7.13}$$

Here γ is a constant, known from routine calibrations used for other growths on this system[5].

Combining Eqs. (7.10) and (7.13), we can use the ion gauge current to measure the effective temperature response of the cell. The result is

$$T_{\text{eff}}(t) = -\beta \cdot \left[W_{-1}\left(-\frac{\beta\gamma}{\alpha} j_{\text{IG}}(t) \right) \right]^{-1}. \tag{7.14}$$

Fig. 7.1 shows the T_{eff} step response of our Al cell, measured in this way. We see that T_{eff} lags, overshoots, and then settles back to T_{in}. From this data, we could estimate the transfer function directly as $H(f) \approx$

[5]Here again, we are using T_{eff} in place of the temperature of the emitting surface. As before, this is not quite correct, but the effect will be mitigated as long as the same approximation is used when calibrating γ.

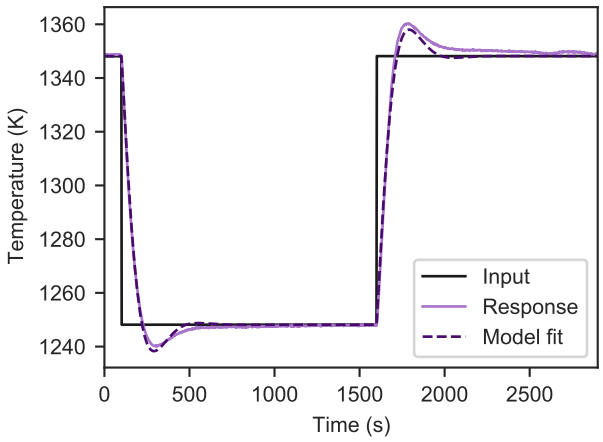

Figure 7.1: Step test of our Al cell (growth number G0452 in the QNC-MBE database). Black: the input temperature to the PID controller. Purple, solid: the Al flux response, measured by ion gauge, and then converted into an effective temperature using Eq. (7.14). Purple, dashed: response predicted by the second-order linear model (Eq. (7.15)) after fitting the parameters K, ζ, and τ with least-squares.

$T_{\text{eff}}(f)/T_{\text{in}}(f)$, but it is clear that this will incorporate noise and other errors into the transfer function. Instead, we extract the key dynamics by restricting the transfer function to a simpler, second-order form,

$$H(f) = \frac{T_{\text{eff}}(f)}{T_{\text{in}}(f)} \approx \frac{K\zeta \cdot (i2\pi f\tau) + 1}{(i2\pi f\tau)^2 + 2\zeta \cdot (i2\pi f\tau) + 1}, \tag{7.15}$$

where K, ζ, and τ are parameters to be fitted. Depending on the particular effusion cell, another form of the transfer function might be more useful. In general, the approach is to use just enough poles and zeros to capture the key dynamics. In this case, we have also chosen the form of $H(f)$ so that $H(0) = 1$, which enforces the constraint that T_{eff} should be equal to T_{in} in the steady-state. For more detailed comments on choosing the form of the transfer function, see Appendix B.3.

Furthermore, it should be noted that the transfer function depends on the settings of the cell's control system. Thus, for example, if the PID settings are adjusted, then the transfer function parameters will need to be remeasured. In principle, tuning the control system does not improve the maximum rate of flux change achievable with this method, since that is determined by the thermal mass of the cell itself. However, if the control system is poorly tuned, then the operator may need to request extreme values of T_{in} to achieve a rapid flux change.

The parameters K, ζ, and τ are fitted to give a step response as close as possible to the measured one. The results of this fit for our system are plotted alongside the measured data in Fig. 7.1. Clearly, we see that there are some dynamics here that are not captured by the linear model. This is not surprising, since the step test is an extreme example, and the effusion cell thermal dynamics are known to be non-linear. However, we will see later that this transfer function already provides reasonably good results, and that the missing dynamics can be well-compensated by an iterative correction.

7.5 Shutter transient compensation

The transfer function of Sec. 7.4 describes the response of the cell flux to changes in input temperature, but it does not account for the response due to the cell shutter status. Opening or closing the cell shutter changes the distribution of heat within the cell, which leads to transient variations in flux whenever the shutter is actuated. Often, growths with continuous alloys might also require sharp interfaces, which means we need a general approach that can account for the thermal dynamics due to both changes in input temperature and those due to shutter actuation.

The shutter action can be thought of as inducing a transient in the effective cell temperature, T_{eff}. So, to compensate for the shutter action, we extend our linear model of Eq. (7.12) to

$$T_{\text{eff}}(f) = H(f)T_{\text{in}}(f) + G(f)S(f). \tag{7.16}$$

Here $S(t)$ is the shutter state, which is either 0 (open) or 1 (closed), and $S(f)$ is its Fourier transform. $G(f)$, then, is the transfer function which describes the shutter response. (Note that since S is dimensionless, G has units of temperature.)

The shutter response is measured using the same approach as the temperature input response. We perform a "step" test by opening the shutter (stepping $S(t)$ from $1 \to 0$) while keeping T_{in} constant. Meanwhile, we measure the flux response with an ion gauge, and convert it into an effective temperature response $(T_{\text{eff}}(t) - T_{\text{in}})$. We repeat this test multiple times at different temperature setpoints and calculate the average shutter transient. Then, we fit the shutter transfer function to the following form:

$$G(f) \approx \frac{K_s\zeta_s \cdot (i2\pi f\,\tau_s) + \eta_s}{(i2\pi f\,\tau_s)^2 + 2\zeta_s \cdot (i2\pi f\,\tau_s) + 1}. \tag{7.17}$$

Here we have four parameters to fit – K_s, ζ_s, τ_s, and η_s – and the results of this fit are shown in Fig. 7.2. As with the thermal transfer function, $H(f)$, the exact form of $G(f)$ may need to be adjusted, depending on the particular cell and shutter system. Here again, we can gain some intuition by examining the steady-state behaviour. In this case, we have $G(0) = \eta_s$. So, in the steady state, Eq. (7.16) becomes simply $T_{\text{eff}} = T_{\text{in}}$ if the shutter is open and $T_{\text{eff}} = T_{\text{in}} + \eta_s$ if the shutter is closed. Thus, the parameter η_s can be interpreted as the net increase in effective cell temperature after closing the shutter (i.e., the difference between the initial and final values in Fig. 7.2).

As with the thermal transfer function of Sec. 7.4, this shutter transfer function, $G(f)$, is only an approximation of much more complicated non-linear dynamics. For example, this model assumes that the shutter response is independent of the input temperature, and that the T_{eff} response upon closing the shutter is just the negative of the response upon opening – neither of which is necessarily the case. Nevertheless, it provides a relatively simple approximation of the shutter transient, which, if necessary, can be refined using the iterative correction scheme described later. Such an approach is not only useful for growths incorporating continuous alloys: it could readily be applied to mitigate the shutter transients in almost any growth where precision is paramount.

7.6 Mock growth

To test these dynamical models, we applied them to a real structure – a periodic sequence of parabolic quantum wells (PQWs). These wells are grown in $\text{Al}_x\,\text{Ga}_{1-x}\,\text{As}$ with composition, x, ranging from 2–30 %, and are designed to give an intersubband transition energy of around 12.4 meV (which corresponds to

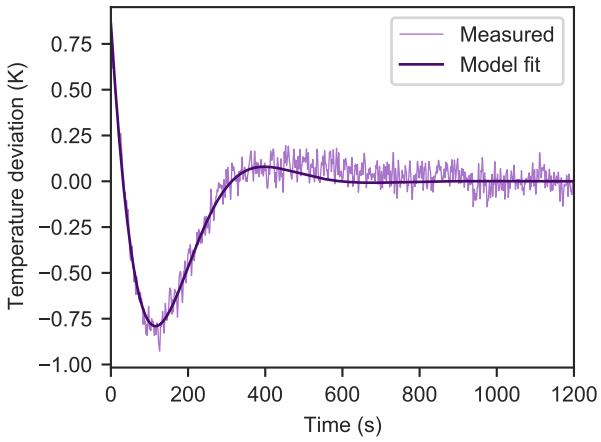

Figure 7.2: Shutter transient response of our Al cell (growth number G0452 in the QNC-MBE database). Plotted here is the deviation of the effective temperature from its steady state value, after opening the shutter at $t = 0\,\mathrm{s}$. Light purple: Al flux response, measured by ion gauge and then converted into an effective temperature using Eq. (7.14). Dark purple: response predicted by the second-order linear model (Eq. (7.17)) after fitting the parameters K_s, ζ_s, τ_s, and η_s.

3.0 THz). In between the smooth parabolic wells, we place 2 nm rectangular GaAs wells, which are created by closing the Al shutter. These wells are where the Si dopants are placed for modulation-doped PQWs [60, 130]. We assume a constant Ga flux such that the growth rate of pure GaAs would be 0.15 nm/s.

Using the transfer functions $G(f)$ and $H(f)$, measured above, we numerically invert Eq. (7.16) to find the Al cell input, $T_{\mathrm{in}}(t)$, required to grow a periodic sequence of these PQWs.

Since it is difficult to measure the composition accurately during a real growth, we perform a "mock" growth – applying the calculated Al cell input (shown in Fig. 7.3) to the Al cell, and without supplying any Ga or As flux. We measure the Al flux – again, using an ion gauge placed in front of the substrate holder – and then calculate what the composition would have been if it were an actual growth with Ga flux. We repeat the test under periodic conditions by applying the input 10 times in a row.

Fig. 7.4 shows representative results for this mock growth. The measured flux, converted to an effective composition is quite close to the target structure, with a root-mean-square composition error of only $\Delta x = \pm 0.70\,\%\mathrm{Al}$. We see that most of the error is concentrated at the tops of the barriers, where the Al cell temperature must change rapidly. Depending on the demands of the structure, this level of precision may already be sufficient, and could perhaps be improved further by optimizing the underlying PID parameters.

7.7 Iterative improvement

One notable feature of the mock growth results in Fig. 7.4 is that the position-dependent composition error is nearly the same for each well. This suggests that most of the error does not come from random instabilities in the effusion cell, but rather from predictable dynamics that have not been captured by our

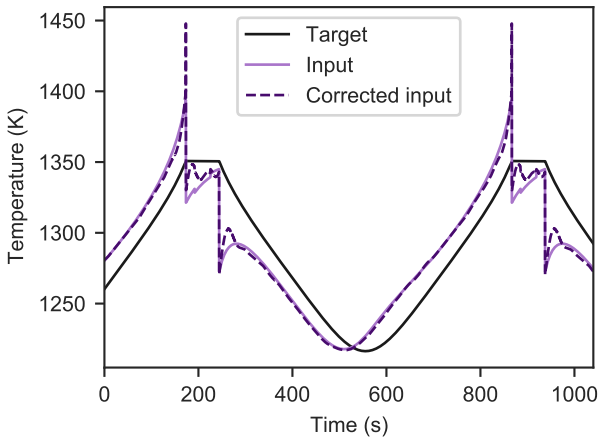

Figure 7.3: Black, solid: calculated effective temperature, $T_{\text{eff}}(t)$, required to grow parabolic quantum wells for the mock growth outlined in Sec. 7.6. Light purple, solid: input temperature expected to generate this target effective temperature, as calculated by Eq. (7.16). This was the input used for the mock growth in Sec. 7.6 (growth number G0472 in the QNC-MBE database). Dark purple, dashed: corrected input, calculated using the iterative method outlined in Sec. 7.7. This was the input used for the mock growth in Sec. 7.7 (growth number G0475 in the QNC-MBE database). In both cases, the input is generated with periodic boundary conditions.

simple linear model. So, in principle, it should be possible to eliminate this error.

To do this, we develop a simple iterative correction scheme. For a given target structure, we begin with a mock growth similar to Sec. 7.6. An initial cell input, $T_{\text{in}}^{(0)}(t)$, is calculated using Eq. (7.16), and the resulting flux is measured and then converted into an effective temperature, $T_{\text{eff}}^{(0)}(t)$. Ideally, this measured temperature would already be equal to the target, $T_{\text{eff}}^{(\text{target})}(t)$, but in practice there will be some error, equal to

$$T_{\text{error}}^{(0)}(t) = T_{\text{eff}}^{(0)}(t) - T_{\text{eff}}^{(\text{target})}(t). \tag{7.18}$$

We would thus like to apply some correction to the input temperature to negate this error. Since we know that our transfer function $H(f)$ is a decent approximation to the dynamics, a reasonable guess for the correction is: $-T_{\text{error}}^{(0)}(f)/H(f)$. That is, the new corrected input temperature would be (in the frequency domain)

$$T_{\text{in}}^{(1)}(f) = T_{\text{in}}^{(0)}(f) - \frac{T_{\text{error}}^{(0)}(f)}{H(f)}. \tag{7.19}$$

This correction scheme could be easily iterated if necessary:

$$T_{\text{in}}^{(n+1)}(f) = T_{\text{in}}^{(n)}(f) - \frac{T_{\text{error}}^{(n)}(f)}{H(f)}. \tag{7.20}$$

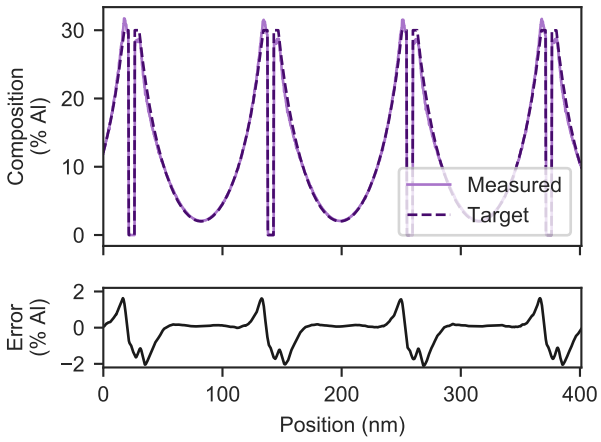

Figure 7.4: Results of the mock growth described in Sec. 7.6 (growth number G0472 in the QNC-MBE database). The top pane shows the target composition profile (dark purple, dashed) and the measured composition (light purple, solid) during the mock growth. The Al flux was measured by ion gauge and then converted into an equivalent composition. The bottom pane shows the difference between the target and measured curves. For clarity, this error curve has been low-pass filtered to remove ion gauge noise and numerical artefacts at the sharp well boundaries, emphasizing the error which comes from the dynamical model. (The measurement in the top pane remains unfiltered.) The full test included 10 wells, but only a representative sample (wells 6, 7, and 8) is shown, for simplicity.

We tested this approach on the parabolic wells of Sec. 7.6. The results of applying this corrected input are shown in Fig. 7.5. After just one corrective iteration, the root-mean-square composition error is reduced from ±0.70 %Al down to ±0.18 %Al. In particular, we see that the composition at the top of the high-composition barriers matches the target much more closely. While another iteration could be applied, the error is already approaching the limits in accuracy of the ion gauge. So this correction scheme provides a simple way to achieve excellent composition tracking.

This iterative method has the downside that it requires at least one additional flux measurement before the growth of any new structure. Fortunately, for periodic structures like superlattices, only a few periods need to be measured, so the flux measurement can be much shorter than the actual growth time. In addition, we found that the Al cell dynamics do not change too dramatically over the course of a growth campaign, even as the cell is depleted. This means that the calibration and mock growth measurements described here can often be re-used for future growths. As a test of this, we repeated the mock growth of Fig. 7.5 about two months later, after growing at least 50 µm of AlGaAs material for other structures. Then, using the exact same temperature input (only corrected for routinely measured changes in α and β – we did not do any new step tests or iterative corrections), we found that the error increased only by a small amount, from ±0.18 %Al to ±0.22 %Al.

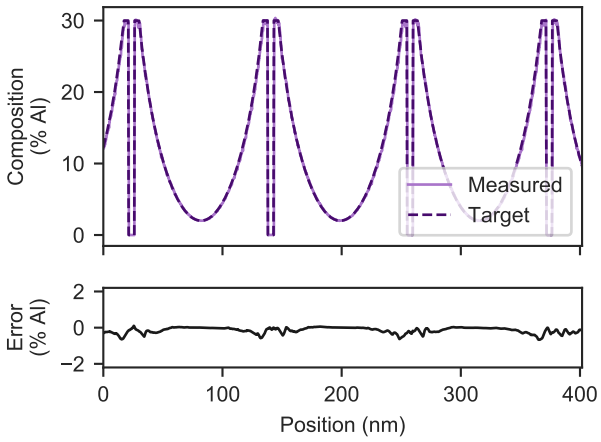

Figure 7.5: Results of the mock growth described in Sec. 7.7, using a corrected input sequence (growth number G0475 in the QNC-MBE database). As in Fig. 7.4, the top pane shows the target and measured Al fluxes, converted into effective compositions. The bottom pane shows the difference between the two. Again, for clarity, the error curve has been filtered, and only a representative sample of the data is shown.

7.8 Characterization with X-ray diffractometry

The beam-flux monitoring ion gauge used for the above tests is convenient, as it gives a full profile of how the Al flux varies with time. However, these measurements rely on an accurate calibration to convert the ion gauge current (pA) into particle flux ($\mathrm{cm}^{-2}\mathrm{s}^{-1}$). (I.e., the γ coefficient discussed in Ch. 6) While such calibrations are routinely used, it is important to grow actual structures to confirm that these "mock" measurements can successfully translate into real PQWs.

To test this, we use an Al temperature input similar to the one calculated above to grow a stack of 10 PQWs. (Specifically, this is the input used in Ref. [115].) A high-resolution x-ray diffractometry (HR-XRD) measurement is performed on the sample in ω–2θ coupled mode, and the results are compared to a simulation of the target structure. Fig. 7.6 shows the HR-XRD results – the fit to the target structure is excellent without any adjustments or fitting parameters. Although HR-XRD is not the most sensitive to distortions in the composition profile, this is a reassuring piece of evidence that the mock-growth flux measurements can be used to grow a real structure successfully.

7.9 Characterization via reflectivity

As another way to verify the accuracy of the BFM for these flux measurements, we did a test growth using in-situ reflectivity. Reflectivity can be used to measure the growth rate of certain layers in real-time [131]. The basic idea is that the reflectivity of a substrate is sensitive to the thickness of layers grown on top of it. This is due to reflection and refraction at the interfaces which give rise to interference patterns that can enhance or suppress the total reflectivity at the surface. As the top layer grows thicker, we can track the change in reflectivity in real-time and fit it to a theoretical model to extract the growth rate.

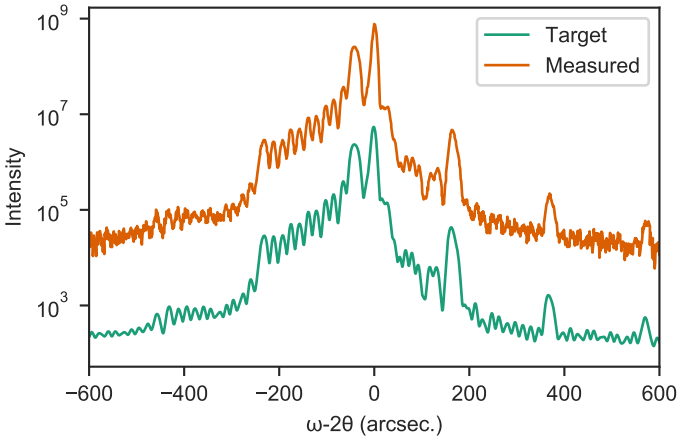

Figure 7.6: HR-XRD data from a grown sequence of 10 PQWs (growth number G0357 in the QNC-MBE database). The black curve shows the expected HR-XRD counts, as calculated by a simulation of the target structure. The red curve shows the actual measured HR-XRD counts, which track the expected curve extremely well.

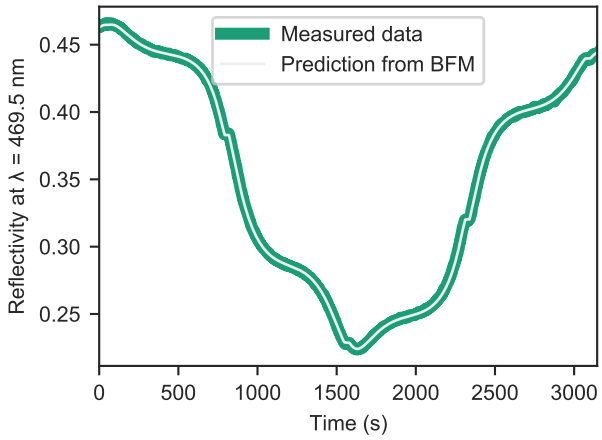

Figure 7.7: Dark (green online): Wafer reflectivity as a function of time during the growth of an AlAs layer on a GaAs substrate (growth number G0636 in the QNC-MBE database). (Several GaAs and AlAs layers had already been deposited before the growth of this AlAs layer.) The reflectivity was measured at a wavelength of 469.5 nm. During the AlAs growth, the Al flux was varied as it would be when growing a sequence of four $Al_x Ga_{1-x} As$ PQWs, so the AlAs growth rate varied as a function of time. Light grey: the predicted reflectivity based on a mock growth measurement with the beam flux-monitoring ion gauge (BFM) ion gauge. The same Al cell input was applied and the flux was measured with a BFM ion gauge.

In this case, we start with a GaAs substrate (potentially with some AlGaAs layers on top) with some effective refractive index, N_s. If we then grow a new layer of thickness L on top, the reflectivity at the surface will become [131]:

$$r(L) = \frac{r_\infty + r_i \exp(-i4\pi NL/\lambda)}{1 + r_\infty r_i \exp(-i4\pi NL/\lambda)} \tag{7.21}$$

where λ is the wavelength at which the reflectivity is being monitored and

$$r_\infty = \frac{1-N}{1+N}; \qquad r_i = \frac{N-N_s}{N+N_s} \tag{7.22}$$

where N is the complex refractive index of the grown layer. Typically, this formula would be used during a constant-rate growth of a single material (i.e., $L = G \cdot t$, where G is the growth rate), which gives rise to nearly sinusoidal oscillations. However, for these continuous alloy growths, we would like to study the reflectivity while the Al flux is changing as a function of time.

In principle, we could monitor reflectivity during the growth of $Al_x Ga_{1-x} As$ PQWs and extract the growth rate. Unfortunately, we cannot do this in practice because we do not have precise enough knowledge of the complex refractive index as a function of composition at the temperatures used for a typical growth. However, if we eliminate the Ga flux, we can grow a pure AlAs layer, for which we *do* have well-calibrated information about the refractive index. As the Al flux varies, the growth rate of the AlAs will vary, leading to a predictable reflectivity as a function of time. So, we can do the following:

1. Apply a time-varying Al input for generating PQWs, as discussed above.
2. Measure the resulting flux using the BFM.
3. Apply the same input, this time along with an As flux to grow a layer of AlAs on GaAs, with the growth rate varying as a function of time.
4. Use the BFM data to predict what the reflectance versus time should be, and compare it to the measured reflectivity data.

This is what was done in Fig. 7.7, and we can see that the predicted reflectivity from the BFM matches very well with the measured reflectivity.

It should be noted that some least-squares fitting had to be done here. The layer underneath the AlAs is not pure GaAs, and so its effective refractive index, N_s, is unknown and must be extracted by fitting [131]. Furthermore, we also allowed the BFM calibration coefficient γ to be fit by the routine, to eliminate the effect of overall calibration errors. (Only a single, constant value of γ was used for the entire fit, though.) This test thus gives good evidence that – as long as the BFM is carefully calibrated with an accurate γ coefficient – it is a reliable flux gauge, whose results can be translated into real growth rates.

7.10 Conclusions

We have developed a method for controlling effusion cell flux which, in our example case of growing continuously graded $Al_x Ga_{1-x} As$, achieved an accuracy approaching the limits of existing measurement techniques. The typical exponential relationship between effusion cell flux and input temperature is

supplemented with a linear dynamical model to account for the transient behaviours of the effusion cell. A simple, second-order model is used, with parameters calibrated by a single step test.

Once calibrated, the linear model can be used to generate arbitrary composition profiles, as long as the effusion cell is capable of changing its temperature quickly enough. The linear model may already provide good enough compositional accuracy for many applications. However, for cases where better control is needed, we have introduced an iterative correction scheme to eliminate any lingering error. With just a single extra calibration measurement for each particular structure, the error can be reduced significantly.

Using this iterative correction, we controlled the composition of an $Al_x Ga_{1-x} As$ PQW sequence to within ±0.18 %Al at standard growth rates above 0.15 nm/s. The method can easily be applied to other composition profiles, and it does not require any modifications to the existing MBE hardware. While we studied the case of continuously graded $Al_x Ga_{1-x} As$ here, it should be re-emphasized that this technique is more about effusion cell control, and it could be applied in other situations as well. This could include other material systems – perhaps even more difficult mixed group V systems like InAs/AlSb. It could also include the mitigation of shutter transients, particularly in highly demanding growths like THz quantum cascade lasers, for which the additional calibration time may be well worth it. With the combination of a simple dynamical model plus iterative correction, we obtain a powerful technique for controlling cell fluxes during MBE growth.

Chapter 8

Room temperature absorption with PQWs

Note: significant portions of this chapter are derived from our publication, Ref. [130], with permission.

In the last chapter, we outlined a technique to grow continuously graded structures with molecular beam epitaxy (MBE), with a particular focus on the example of parabolic quantum wells (PQWs). In this chapter, we will explore the results of applying this technique more thoroughly, with an eye towards terahertz (THz) strong coupling applications. As discussed in detail in earlier chapters, one of the key difficulties in this spectral range is to achieve high-quality THz resonances, particularly above cryogenic temperatures. Intuitively, this can be understood as coming from the fact that a THz resonance requires a small intersubband transition energy, on the order of 10 meV. Thus, at room temperature for example, the thermal energy ($k_B T \sim 25$ meV) will be enough to populate several subbands. In square wells, this leads to an undesirable splitting of the absorption resonance as the temperature rises – on top of linewidth broadening due to the usual scattering mechanisms [105] – which makes it very difficult to achieve strong coupling above cryogenic temperatures [72].

Prior to this work, it had already been shown that PQWs are able to maintain a THz absorption resonance up to room temperature and achieve ultra-strong coupling [60]. However, digitally alloyed PQWs suffer from broad absorption lines – for example, Ref. [67] saw absorption linewidths on the order of 30 % of the ∼3.5 THz centre frequency, with no improvement at low temperature. While this is sufficient for a demonstration of strong coupling, better linewidths are needed for other geometries and devices – in particular, for the metal-insulator-metal (MIM) geometries that would be desirable for polariton scattering studies [75].

In this chapter, we use continuously graded PQWs to demonstrate intersubband (ISB) absorption at 3 THz, with remarkably small linewidths. At low temperatures near 10 K, we see linewidths below 6 % of the transition frequency – a serious improvement over digital PQWs, and even an improvement compared to square quantum wells. Importantly, this narrowness is largely maintained even up to 150 K, well above the temperature of liquid nitrogen. Thus, comparable experiments on square quantum wells that had to be performed with liquid helium could be performed with much more readily available liquid nitrogen using continuously graded PQWs.

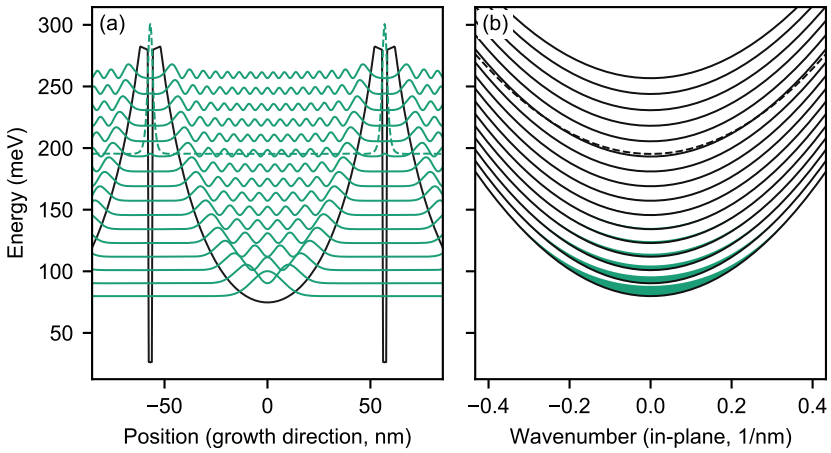

Figure 8.1: (a) Schrödinger-Poisson simulation of a periodic stack of PQWs at 300 K, performed with nextnano++. The bare (undoped) PQW intersubband transition energy is 3 THz. Modulation doping of $1 \times 10^{11}\,\mathrm{cm}^{-2}$ per well is placed in 2 nm square wells in between the PQWs. The black curve shows the conduction band edge, while the grey (green online) curves show the squared envelope wavefunctions, shifted according to their subband energy. Energy is relative to the Fermi level. (b) In-plane subband dispersions of the same simulation. Shaded regions represent the relative electron occupation. In both (a) and (b), the dashed lines correspond to the bound state of the square doping well.

Figure 8.2: A basic schematic of the multipass experiment setup. Light enters through a 45° facet, and bounces through the sample until it reaches the exit facet. Three key regions of the sample are shown, from bottom to top (not to scale): the substrate (blue online), the active region (purple online), and the Ti/Au coating (gold online). Each time the light passes through the active region, a portion of it is absorbed. The sample should be designed so that the light reflects completely at the bottom of the substrate, e.g., by total internal reflection.

8.1 THz absorption in parabolic quantum wells

To study the absorption properties of the continuously graded PQWs, we grew two arrays of PQWs – one with 54 PQWs n-type modulation doped at 1×10^{11} cm^{-2} per well, and the other with 18 PQWs doped at 3×10^{11} cm^{-2} per well. (Growth numbers G0490 and G0489, respectively, in the QNC-MBE database.) The wells were grown in $Al_x Ga_{1-x} As$ using the techniques of Chapter 7, with Al composition continuously varied along the growth direction (illustration in Fig. 8.1). At the bottom of each PQW, the minimum Al composition, x, was chosen to be 2 %, and the barriers were truncated at a maximum composition of 30 %. For these growths, the Al composition profile was chosen to give a parabolic conduction band edge profile based on the bandgap vs composition dependence of Ref. [95]. The PQWs were each 103.8 nm wide, separated by 10 nm barriers comprising 4/2/4 nm of $Al_{0.3} Ga_{0.7} As/GaAs/Al_{0.3} Ga_{0.7} As$. Si delta-doping was placed in the GaAs region in the centre of the barrier.

To analyze the quality of the PQW stacks, we measured the THz transmission with a Fourier transform infrared (FTIR) spectrometer. (A simple schematic is shown in Fig. 8.2.) The samples were surface coated with Ti/Au and shaped into multi-pass waveguides with facets polished at 45°. The 18-well sample was 0.56 mm × 2.48 mm, and the 54-well sample was 0.54 mm × 2.25 mm. Here, the first number is the thickness in the growth direction, and the second number is the length of the waveguide perpendicular to the growth direction and in the same plane as the light propagation. The samples were placed into a continuous-flow cryostat inside a Fourier transform infrared spectrometer (Bruker IFS66V), with polarized broadband THz light incident on the sample. The transmitted spectra were collected with a liquid-He-cooled Si bolometer. For measurements at room temperature (approx. 293 K), the cryostat was not used, to increase the signal-to-noise ratio. To isolate the ISB absorption of the PQWs, ratios were taken between the transmission spectra for TE- and TM-polarized incident light. Because of the ISB transition selection rule, PQWs absorb only the TM-polarized light.

The results are reported in Fig. 8.3. An absorption resonance is observed in both samples near 3 THz; it is already very clear at room temperature and narrows further as the samples are cooled. The absorption

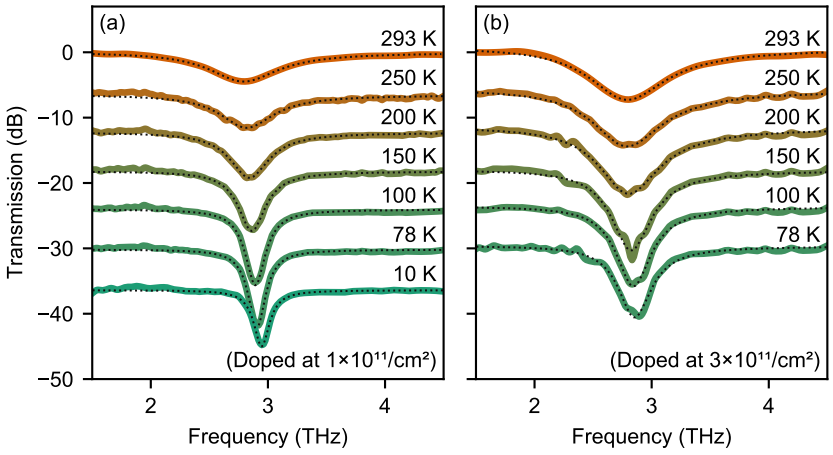

Figure 8.3: Transmission measurements for the two PQW samples over a range of temperatures. (a) 54 PQWs modulation doped at 1×10^{11} cm^{-2} per well (growth number G0490 in the QNC-MBE database). (b) 18 PQWs modulation doped at 3×10^{11} cm^{-2} per well (growth number G0489 in the QNC-MBE database). In both (a) and (b), solid lines show the measured transmission spectra, and dotted lines show the empirical fit. Measurements at different temperatures are offset by 6 dB for clarity.

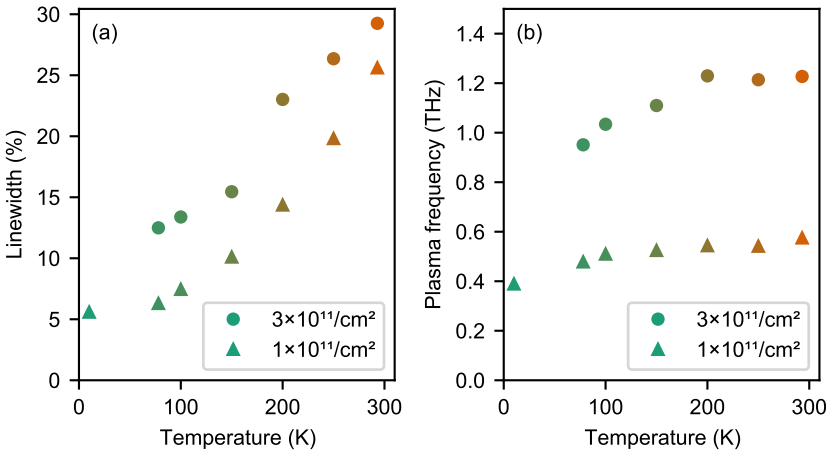

Figure 8.4: (a) Relative linewidth γ/ω_0 as a function of temperature. The linewidth was extracted from fits of the transmission data, and it is expressed as a percentage of the fitted centre frequency. (b) Measured plasma frequency, ω_P of the doped samples, extracted by fitting the data with Eq. (8.1).

behaviour is consistent with a strong, unified quantum well (QW) resonance and minimal inhomogeneous broadening. To extract more quantitative information, we fitted the data with an empirical model. We used a single-mode form of the dielectric tensor for the PQW stack, with growth-direction (z) component as [61]:

$$\frac{1}{\epsilon_{zz}(\omega)} = \frac{1}{\epsilon_s}\left(1 + \frac{\omega_P^2}{\omega^2 - \omega_0^2 + i\gamma\omega}\right) \tag{8.1}$$

Here ϵ_s ($\approx \epsilon_{xx} = \epsilon_{yy}$) is the average static dielectric constant of the PQW stack. The parameters ω_0, γ, and ω_P can be intuitively understood as the resonant frequency, homogeneous linewidth, and plasma frequency, respectively, but they are treated here as parameters to be fitted. This dielectric tensor is fed into a transfer matrix-style model to calculate the power transmission through the sample. In the limits of long wavelength (thin active region) and low absorption, this approach reduces to a simpler exponentially decaying power rule [90]. While the exponential decay model is commonly employed in multipass absorption, here a full consideration of the boundary conditions at the interfaces is necessary to explain the apparent asymmetry of the peaks, which arises due to variations in the electric field overlap with the PQW stack as a function of wavelength. The mathematical details and derivation of the model used here are provided in Appendix C.

The experimentally extracted linewidths, γ, are shown in Fig. 8.4(a) as a function of temperature. They are consistently smaller than 30 % of the transition frequency, reaching just 5.6 % for the lighter-doped, 54-well sample at 10 K. Remarkably, for this sample, γ remains below 11 % right up to 150 K, while the absorption strength – as measured by the experimentally extracted plasma frequency, ω_P[1] – is nearly constant with temperature (Fig. 8.4(b)). This is in contrast to THz transitions in square QWs, which experience both significant broadening and a sharp drop-off of the fundamental absorption peak with temperature. Of particular interest for experimental applications is the region 77–150 K, wherein these PQWs exhibit linewidths better than 11 %, providing a high-quality THz resonance above liquid nitrogen temperatures.

Another notable feature of the absorption data is the stability of the resonant frequency, ω_0, which varies only from 2.8–3.0 THz across all the temperatures and doping levels that we have studied. The stability of the peak position with doping, in particular, provides an elegant demonstration of the results of Ref. [108], similar to the Kohn theorem for cyclotron systems.

8.2 Modelling the absorption results

As we have seen in Ch. 5, understanding this and more fully modelling the absorption behaviour of these PQWs can be a subtle task. At the doping levels studied here, the Hartree potential leads to a significant redshift of the ISB transition energy, on the order of several meV. Furthermore, over most of the temperatures and doping levels studied, several subbands are occupied, which would seemingly preclude the use of a single-resonator model like Eq. (8.1).

To explain the simplicity and stability of our observed absorption peaks, we must go beyond the single-electron approximation and include interactions between the different ISB plasmon modes. In

[1] In cases where the absorption coefficient is well-defined, the integrated absorption is proportional to ω_P^2, and does not depend on γ or ω_0: see Appendix C.4. This suggests ω_P as a reasonable measure of absorption strength, and it has the advantage of being well-defined here, even when the absorption coefficient is not.

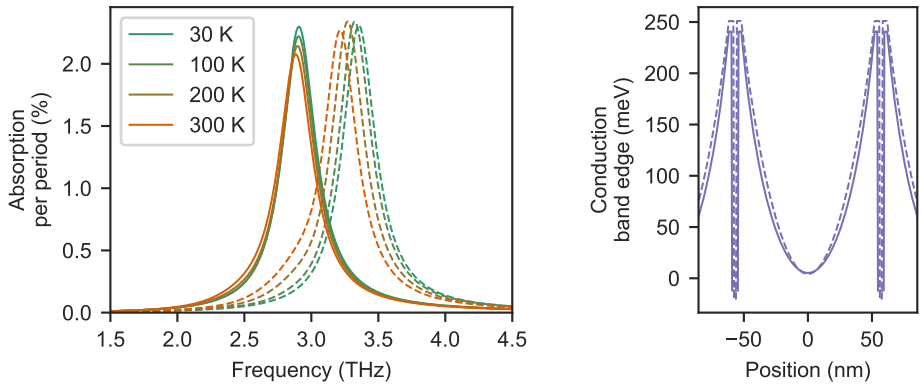

Figure 8.5: Left: Calculated absorption spectra for the sample doped at 1×10^{11} cm^{-2}. The absorption was calculated using the semiclassical model outlined in Ch. 5. A constant, arbitrarily chosen linewidth, $\gamma = 0.3$ THz was used to isolate the multisubband plasmon effects. The composition profile is assumed to follow that of the target grown sample. However, two different relationships are used for the $Al_x Ga_{1-x} As$ bandgap vs composition. The dashed lines correspond to the cubic dependence recommended by Vurgaftman *et al.* [95]. The solid lines correspond to the linear dependence of Wasilewski *et al.* [118]. In both cases, a 60:40 split was used between conduction and valence band offsets. **Right**: the two conduction band (CB) edge profiles versus position in the growth direction. These are calculated at 30 K in the absence of doping. Again, the dashed line corresponds to Vurgaftman *et al.* [95], and the solid line corresponds to Wasilewski *et al.* [118]. Since Vurgaftman *et al.* was used for the initial design of the composition profile, it gives a near-perfect parabola (dashed), while Wasilewski *et al.* suggests that the composition profile would give a distorted parabola.

the effective mass approximation, a QW can still be modelled with an effective permittivity tensor, with a growth-direction component of the form

$$\frac{1}{\epsilon_{zz}(\omega)} = \frac{1}{\epsilon_s}\left(1 + \sum_j \frac{\Omega_{P,j}^2}{\omega^2 - \Omega_j^2 + i\Gamma_j\omega}\right) \tag{8.2}$$

Here $\Omega_{P,j}$, Ω_j, and Γ_j are the weights, resonance frequencies, and homogeneous linewidths, respectively, of the absorption modes. However, these are now collective "multisubband" plasmon modes which arise from the interaction between different ISB transitions [109]. The weights and resonance frequencies can be calculated using either a quantum approach [97], or a semiclassical approach [101], which are equivalent in the long-wavelength limit.

More generally, in the case of non-parabolic subbands, the picture becomes slightly more complicated and Eq. (8.2) is only approximately true, as it becomes difficult to assign a finite set of modes to the system. Nevertheless, as we have discussed in Ch. 5, it is still possible to calculate an effective dielectric tensor using a semiclassical model.

In the case of a PQW, quantum modelling in the effective mass approximation predicts that almost all the absorption strength should be concentrated in a single mode j, whose frequency, Ω_j, is locked to the bare ISB transition of the undoped well – regardless of doping and temperature [109]. More sophisticated

modelling with non-parabolic subbands confirms the same idea (see Sec. 5.7): PQWs have a robust absorption peak over a wide range of dopings and temperatures. This notable behaviour of harmonic systems is exactly what we see experimentally in Fig. 8.3.

While the overall behaviour is as expected, a more precise comparison between model and theory illuminates lingering uncertainties. The composition profile of the grown structures was designed using nextnano++ [91] – a Schrödinger-Poisson solver – to give a perfectly quadratic conduction band in the absence of doping. The full model (quantum or semiclassical) then predicts that this composition profile should give a stable absorption resonance at around 3.3 THz – shown for the 54-well sample in the dashed lines of Fig. 8.5.

This predicted absorption frequency is around 0.3 THz higher than what is experimentally observed. One possible explanation for this discrepancy is the dependence of the $Al_x Ga_{1-x} As$ bandgap on composition. Nextnano++ uses, by default, the cubic dependence recommended by Vurgaftman *et al.* [95], which has been used quite widely in the community as the go-to source for III-V energy band parameters. However, careful measurements of $Al_x Ga_{1-x} As$ composition and bandgap show that a linear dependence would be more appropriate for $x < 0.45$ [118]. (Vurgaftman *et al.* in fact acknowledge this within in the text, although they ultimately recommend a non-linear dependence – presumably to give a "best fit" over the entire composition range $0 \leq x \leq 1$.) This distinction amounts to a minor correction for QWs with single-composition barriers but could have a significant impact on smoothly graded PQWs, as clarified in Fig. 8.5. A relatively small deviation in the conduction band edge (Fig. 8.5, right) widens the PQW and leads to a significant shift in the peak position. We see that the bandgap dependence of Ref. [118] predicts peak positions around 2.9 THz, which is more in line with the experimentally measured peaks 2.8–3.0 THz.

8.3 Photoluminescence excitation on undoped samples

This issue of the conduction band edge dependence on composition should appear whether doping is present or not. So, to further characterize the PQWs, we performed PLE measurements on undoped samples at ∼10 K. In this case, a stack of 10 undoped PQWs was grown using otherwise identical growth conditions as the doped samples (growth number G0492 in the QNC-MBE database). The 2 nm GaAs wells were skipped so that the barriers were 8 nm of $Al_{0.30} Ga_{0.70} As$. The results are shown in Fig. 8.6.

For PQWs like these, we expect to see an evenly-spaced ladder of subbands in both the conduction band (CB) and heavy hole band (HH). As such, the PLE peaks are expected to obey

$$E_{\mathrm{CB}_n\mathrm{HH}_m} - E_{\mathrm{CB}_1\mathrm{HH}_1} \approx (n-1)\Delta E_{\mathrm{CB}} + (m-1)\Delta E_{\mathrm{HH}} - E_{mn}^{\mathrm{exc}} \tag{8.3}$$

where $E_{\mathrm{CB}_n\mathrm{HH}_m}$ is the position of the PLE absorption peak for a transition from the m-th heavy hole subband to the n-th conduction subband, ΔE_{CB} and ΔE_{HH} are the conduction and heavy hole subband spacings respectively, and E_{mn}^{exc} is the exciton binding energy of the transition. Here, E_{mn}^{exc} is calculated with a variational approach [132], using envelope wavefunctions calculated with nextnano++'s 8-band $\mathbf{k}\cdot\mathbf{p}$ method.

The subband spacings, ΔE_{CB} and ΔE_{HH}, were extracted by fitting Eq. (8.3) to the peak positions in Fig. 8.6. This required a correct labelling of the PLE peaks, which was not immediately apparent in certain cases. The CB_3HH_1 and CB_2HH_4 peaks are expected to appear close together, as are the CB_4HH_2 and CB_3HH_5 peaks. In these two cases, there was no simple selection rule to determine a priori which of the two peaks should dominate. To ensure that this did not impact the results, we allowed for the possibility that both peaks were present and superimposed on top of each other. In the end, a good fit could only be obtained under the assumption that the CB_2HH_4 and CB_3HH_5 peaks dominated.

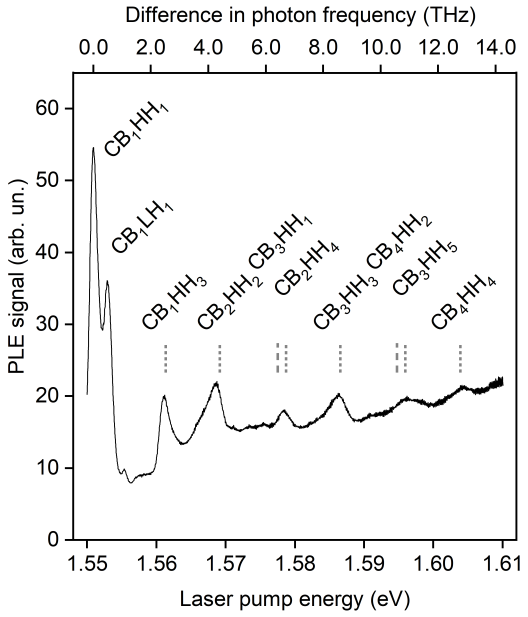

Figure 8.6: Photoluminescence excitation (PLE) measurements at ∼10 K from the undoped PQW stack (growth number G0492 in the QNC-MBE database). The sample was pumped at 60° with TE-polarized light from a tunable Ti:sapphire laser. At ∼10 K, photoluminescence (PL) of the undoped PQWs is primarily due to the CB_1HH_1 and CB_1LH_1 exciton peaks around 1.55 eV. To generate the PLE spectrum, we measured the total intensity of these PL emissions (vertical axis) while varying the frequency of the pump (bottom horizontal axis). The top horizontal axis shows the difference in photon frequency relative to the fundamental CB_1HH_1 peak. The text symbols of CB and HH refer to conduction band and heavy hole band. Dotted and dashed lines with text labels correspond to empirically predicted PLE peak positions, based on the assumption of equally spaced subbands. Thicker dashed lines indicate peaks which were rejected based on the fitting (as discussed in the text).

The subband spacings extracted by this fitting method were $\Delta E_{\mathrm{HH}}/h = 1.1\,\mathrm{THz}$ and $\Delta E_{\mathrm{CB}}/h = 3.0\,\mathrm{THz}$. In comparison, 8-band $\mathbf{k}\cdot\mathbf{p}$ modelling (using the bandgap dependence of Ref. [118]) predicted spacings of $\Delta E_{\mathrm{HH}}/h = 1.1\,\mathrm{THz}$ and $\Delta E_{\mathrm{CB}}/h = 2.9\,\mathrm{THz}$. The discrepancy here is within 4 %, which is quite good considering that the spacing predicted by $\mathbf{k}\cdot\mathbf{p}$ is quite sensitive to uncertainties in the $\mathrm{Al}_x\,\mathrm{Ga}_{1-x}\,\mathrm{As}$ material parameters. Moreover, the calculation predicts that the first few conduction and heavy hole subbands should be equidistant (to within 1 %) for the grown sample. This is a result which is independent of the material parameter uncertainties, justifying the approximations in Eq. (8.3).

It should also be noted that this CB subband spacing extracted from PLE is consistent with the absorption resonances, particularly if we consider the $1 \times 10^{11}\,\mathrm{cm}^{-2}$-doped sample at 10 K (Fig. 8.3(a)). This agreement between the absorption results and the bare ISB frequency calculated by PLE further supports the result of Ref. [108] that the PQW absorption frequency is independent of the number or distribution of electrons. It would, further, seem to suggest that the linear bandgap dependence of Ref. [118] is closer to the correct one – however, other discrepancies still need to be addressed. For example, the absorption resonances based on Ref. [118] (Fig. 8.5, solid) predict no shift with temperature, while Figs. 8.3(a) and 8.3(b) show a clear redshift with increasing temperature.

One possible explanation for this is our imperfect inclusion of the $\mathbf{k}\cdot\mathbf{p}$ dispersions in the calculation of Fig. 8.5, which is based on the semiclassical model of Ch. 5. As discussed in Sec. 5.1.3, it is difficult to account for subband non-parabolicity in the context of multisubband plasmons. This is particularly true in PQWs because the length scales of each transition are different, which prevents [97] the assumptions employed in, e.g., Ref. [100]. Furthermore, for PQWs in $\mathrm{Al}_x\,\mathrm{Ga}_{1-x}\,\mathrm{As}$, the effective mass for each subband is different, as the electrons in higher subbands see a larger aluminum composition on average. Thus is will likely be important to properly account for non-parabolicity in the $\mathbf{k}\cdot\mathbf{p}$ framework, and work on this is underway.

Nevertheless, there is no guarantee that including the full subband dispersions would explain the experimentally observed redshift. Another possibility is that the 60:40 split between conduction and valence band offset is not perfectly constant with composition. As with the bandgap, this might only amount to a minor correction for square wells but could still have a significant impact on PQW behaviour.

Such effects could also impact the absorption linewidth of the grown wells, if they cause the singular multisubband plasmon peak to break down into several peaks at similar frequencies. At present, it is near-impossible to determine if such effects are contributing to the observed linewidths in these samples. Other key factors could include electron-electron scattering, or scattering with ionized impurities, phonons, or the alloy itself [105]. However, the above observation suggests that further reduction of the linewidth – beyond what we already observe in Fig. 8.4(a) – could be possible by further refining the experimentally implemented potential profile. To understand this would require better modelling, incorporating the full $\mathbf{k}\cdot\mathbf{p}$ picture, as well as improved knowledge of key $\mathrm{Al}_x\,\mathrm{Ga}_{1-x}\,\mathrm{As}$ material properties. Other modifications, such as increased barrier width to reduce impurity scattering, could also be beneficial depending on which factors are currently limiting the linewidth.

8.4 Preliminary strong coupling measurements

Although the linewidths of these samples at room temperature are likely too large to perform scattering studies of the type described in Ch. 4, we expect them to be sufficient at lower temperatures. We therefore fabricated the 54 PQW sample into MIM cavity geometries to observe strong coupling. The bottom gold plate was formed by wafer bonding and substrate removal (the samples were grown with an AlGaAs etch stop between the substrate and the active region). The top grating was formed by Ti/Au coating and

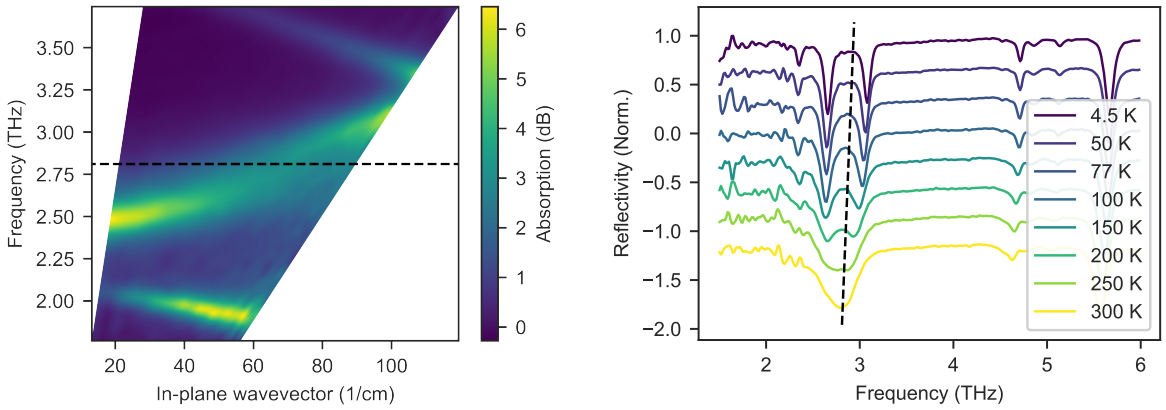

Figure 8.7: Measured reflection from the 54-well PQW sample, growth number G0490, (doped at 1×10^{11} cm^{-2} per well), placed in a MIM cavity with period $\Lambda = 42\,\mu$m and a duty cycle of 85 %. (See Fig. 4.1 for a schematic of a similar geometry.) Light was shined in the plane perpendicular to the slits, transverse magnetic (TM) polarized, at an angle θ relative to the growth axis. The angle θ was converted to an in-plane wave number $k = (\omega/c)\sin\theta$. **Left**: Reflectivity spectrum versus in-plane wave number at room temperature for a sample with grating periodicity $\Lambda = 42\,\mu$m. Reflectivity spectra were measured in a FTIR spectrometer at angles θ ranging from 13° to 73°. The dashed line shows the PQW absorption frequency in the absence of the MIM cavity. **Right**: Reflectivity spectra at a fixed angle $\theta = 15°$ at different temperatures, for the sample with grating period $\Lambda = 36\,\mu$m and a duty cycle of 85 %. Spectra are offset vertically for clarity. The dashed black line again shows the PQW absorption frequency (note that it varies slightly with temperature).

patterned with optical lithography. We formed several gratings with different periodicities, Λ, around 40 µm, each with a duty cycle of 85 %.

An angle-resolved reflectivity spectrum was measured at room temperature and shows the dispersion (Fig. 8.7 left). Even at room temperature, we can see hints of the onset of strong coupling, however there is no visible splitting of the cavity dispersion. This is not surprising, since we know from the modelling of Sec. 4.3 that strong coupling in dispersive MIM geometries can be sensitive to the QW absorption linewidth. For this sample, the linewidth is around 24 % at room temperature. Still it is possible that with improvement to the design of the cavity and PQWs, strong coupling in dispersive MIM cavities at room temperature could be possible.

If we lower the temperature (Fig. 8.7, right) we see that strong coupling and the formation of polariton modes is clearly possible with this sample. At 150 K and below there is a visible Rabi splitting of the absorption peak near 3 THz, with a maximum splitting of 0.42 THz below 50 K. In the future, we hope to perform liquid-nitrogen temperature (78) scattering studies on these samples or similar.

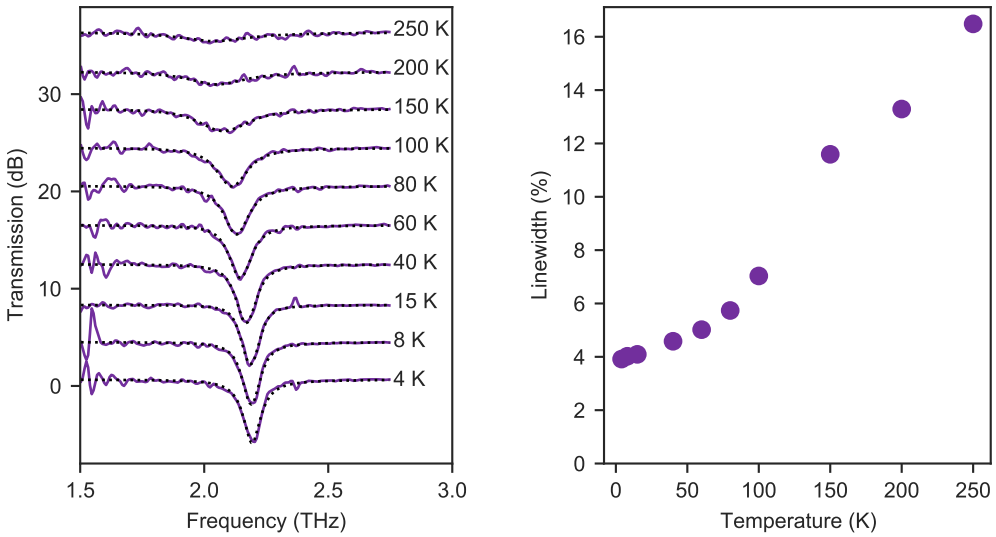

Figure 8.8: Left: Transmission measurements for an 8-well, 2.1 THz PQW sample, using an improved well design (growth number G0643 in the QNC-MBE database). The solid lines show the measured spectra, and dotted lines show the empirical fit to Eq. 8.1. The curves have been offset by 4 dB for clarity. **Right**: Linewidths, γ, (expressed as a percentage of the centre frequency) extracted from the fit.

8.5 Improved parabolic well samples

Later, after the learnings of the sections above, we grew new PQW samples and found that it was possible to achieve even smaller linewidths. These new samples targeted around 2.1 THz, and used only 4–8 wells doped at 1×10^{11} cm^{-2} per well. Unlike the much thicker samples above, which would be more appropriate for the scattering studies discussed in Ch. 4, these thinner active regions are designed to be incorporated in THz "meta-atoms" for optical switching studies [88, 89]. The well design was also changed slightly, compared to the samples above. This time, we assumed a linear relationship between bandgap and $Al_x Ga_{1-x} As$ composition, along the lines of Ref. [118]. We also used a lower barrier composition of $x = 0.2$ and doped straight into the $Al_{0.2} Ga_{0.8} As$ barrier without using doping wells. We used thicker 20 nm barriers, and the wells were 130.8 nm wide.

Because of the small number of wells, the overall absorption strength is significantly reduced. Nevertheless, we were able to see multipass absorption (results for the 8-well sample are shown in Fig. 8.8) up to ~250 K, beyond which the absorption is too weak to see through the noise. For these measurements, we used the same FTIR spectrometer and sample preparation technique as before, but with an improved cryostat which allowed the measurements to be performed under vacuum. This eliminates THz absorption in the air along the beam path, improving the signal-to-noise ratio. This was particularly important for these few-well samples which have a much smaller overall absorption strength.

As with the 3 THz samples, the position of the peak redshifts slightly as the temperature increases. The absorption strength, as quantified by ω_P, is approximately constant with temperature, although the peak absorption drops as the linewidth increases. More interesting, though, is that the linewidths are

reduced even further compared to the 3 THz samples. At 4 K, we see a linewidth of just 3.9 % (which is even smaller in absolute terms, since the centre frequency is lower). The linewidths again remain below 15 % up to 200 K, although the numbers at higher temperatures are less reliable due to the weak signal. It is difficult to say for sure what caused the improvement, but it could be related to our use of a linear bandgap vs. composition dependence in the design. It could also be due to the wider wells and wider barriers, which place the dopant ions further away from the electron wavefunctions and reduce ionized impurity scattering. More study would be needed to elucidate these effects. Nevertheless, it is clear that continuously graded PQWs can provide very high quality absorption resonances at THz frequencies.

8.6 Conclusion

In conclusion, we have demonstrated the growth of high-quality 3 THz harmonic oscillator arrays in $Al_x Ga_{1-x} As$ alloys by faithfully tracking the designed composition profile during MBE growth. The system oscillates at 3 THz, from 10 K up to room temperature, exhibiting narrow linewidths. In line with Ref. [108], the oscillation frequency is largely independent of doping and temperature, despite the presence of significant Hartree potentials. Discrepancies between the model and the results point towards key uncertainties in the understanding of such systems. With small changes to the design, we also demonstrated PQWs around 2.1 THz with even smaller linewidths, down to 3.9 % of the centre frequency at 4 K.

The very high quality of these PQW systems opens up several new vistas. Combined with THz optical cavities, these systems could already enable the implementation of high-quality THz polariton devices operating above liquid-nitrogen temperatures. Even in multipass geometries similar to those employed in this work, PQW stacks with low non-radiative scattering rates could potentially be used to explore regimes in which radiative losses play a more prominent role – analogous to demonstrations of (super)-radiantly limited linewidths [133, 134]. Another very appealing research direction would be the demonstration of ISB transitions in the sub-THz region (200-600 GHz). This would enable the study of the transport of a quantum hall system in strong coupling with a microwave cavity, something that currently is possible only with Landau level systems at extremely low temperatures.

Chapter 9

Half parabolic quantum wells

Throughout this thesis, we have studied parabolic quantum wells (PQWs) in the context of terahertz (THz) polaritonics. In our modelling and design discussions of Chapters 4 and 5, we have seen how PQWs provide a promising alternative to square wells due to their robust THz absorption resonance, even up to room temperature. In Chapters 7 and 8, we examined the growth of PQWs and saw that we can achieve very sharp linewidths by growing PQWs in continuously graded $Al_x Ga_{1-x} As$. One might wonder, though, whether PQWs are the *only* wells that exhibit characteristics that are useful for THz polaritonics.

In this chapter, we will look at one possible alternative: the *half* PQW. These are, essentially, PQWs that have been truncated in the middle. See Fig. 9.1, for an example structure. Half parabolic quantum wells (HPQWs) are interesting because, ideally, they have an equally spaced ladder of subbands exactly like a PQW. However, they lack the same symmetry as PQWs. This could be of interest for the study of non-linear optical phenomena such as polariton scattering [86]. One potential issue with PQWs is that they are effectively supplying a linear restoring force to the electrons, and it is unclear therefore whether they will exhibit strong non-linear behaviour. Much more work is needed on the theoretical side to understand such issues, but it nevertheless means that HPQWs could be an interesting avenue to explore in the context of non-linear phenomena. Furthermore, we can exploit the growth technique developed in Ch. 7 to avoid the use of digital alloys.

9.1 Design challenges for half parabolic wells

Given the near-equally spaced subbands of HPQWs, it is tempting to expect that they will exhibit the same robust absorption as PQWs. Unfortunately, this is not necessarily the case – we must remember that the undoped single-electron Schrödinger picture is only the first step in the calculation (Ch. 5). Doping gives rise to a Hartree potential which can shift the level spacing, and the formation of multisubband plasmon modes can affect the absorption even further. And indeed, it turns out that HPQWs do not always have robust absorption, depending on the doping level and how the doping is applied. In Fig. 9.2 (top panels) we see the absorption profiles for HPQWs modulation and bulk doped at 1×10^{11} cm^{-2} per quantum well (QW). At this doping level, the modulation doped wells exhibits a significant degradation at room temperature, which appears to be from a splitting of the absorption peak. The bulk doped wells, on the other hand, maintain a nearly unified absorption peak up to room temperature.

To gain some intuition why this is the case, we can look at the conduction band and wavefunctions

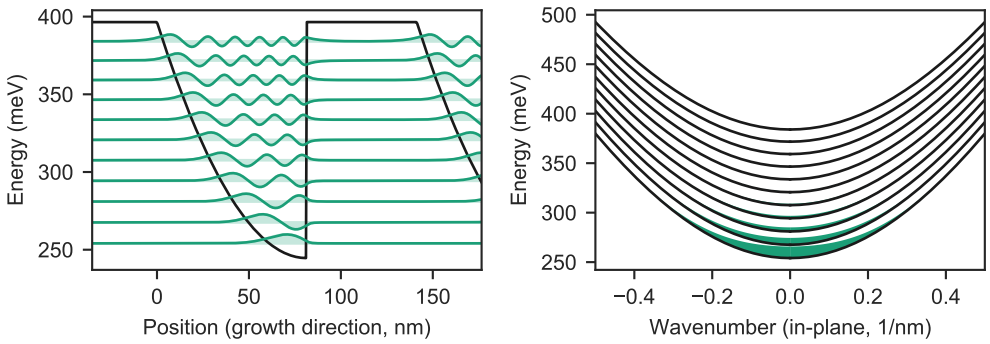

Figure 9.1: Wavefunctions and subband dispersions for a HPQW system. Similar to Fig. 5.1, these results were calculated by Schrödinger-Poisson simulations with nextnano++, with a small 3-band $\mathbf{k} \cdot \mathbf{p}$ correction to the dispersion. The structure comprises $Al_x Ga_{1-x} As$ with composition x ranging from 0.02 to 0.20. The well region is 81.28 nm wide, which gives an absorption frequency around 3.0 THz in the limit of low temperature and low doping. The wells are separated by 60 nm $Al_{0.20} Ga_{0.80} As$ barriers, and periodic boundaries are assumed. For this figure, the well is undoped, and the calculation is performed at 300 K. **Left panels**: the conduction band edge (black) and wavefunctions (green). The vertical scale of the wavefunctions is arbitrary, but they are offset by their corresponding energies. (Energy is measured relative to the Fermi level.) **Right panels**: the subband dispersions (black) as a function of in-plane wave number, $k_{\|}$. The relative level occupation as a function of $k_{\|}$ is shown for each subband by the height of the purple filled regions.

(Fig. 9.2, bottom panels) which have been calculated self-consistently (i.e., including the Hartree potential) for the doped HPQWs. In the modulation doped case, we can see that there is a significant distortion of the half-parabolic potential. This is because the negative electrons in the well are separated from the positive charges in the Si delta, so a significant electric field is created. The bulk doped well, on the other hand, is much less distorted because the electrons and dopants are both located inside the well, so their electric fields cancel to a significant degree. It should be stressed, of course, that this is still not the full picture, and one must include multisubband plasmon effects to determine the actual absorption behaviour – for example, we have seen (Ch. 5.7) that in modulation doped PQWs, band-bending from the Hartree potential does not degrade the absorption quality. However, it is clear in this case that modulation doped HPQWs are *not* robust to charge redistribution effects.

9.2 Bulk doped half parabolic wells

Best on the results of Fig. 9.2, it appears that bulk doping should be the more promising path for HPQWs, when pursuing high-quality THz absorption at room temperature. There is a downside to bulk doping, however, which is not captured by our modelling: ionized impurity scattering. The very Si ions which help to cancel out the potential created by the electrons in the well will also act as scatterers to decrease the non-radiative lifetime of the plasmons [105]. With a large number of ionized impurities in the QW region, we might expect a noticeably broader linewidth, especially as the doping density is increased.

As a preliminary test, we tried growing a stack of 43 HPQWs of the same design described in Fig. 9.1, but bulk doped throughout the well region for a total of 5×10^{10} cm^{-2} per well. The sample was grown with a continuously graded $Al_x Ga_{1-x} As$ alloy, using the technique of Ch. 7. We performed multipass measurements on the sample, using the same setup that was used in Ch. 8. Unfortunately, while the sample did have a THz absorption peak at low temperature, the linewidth was around 18 % at 77 K. Clearly, this is not as good as the PQWs we saw in Ch. 8, which saw linewidths more like 6 % at this temperature. There are several potential reasons for this, but certainly one candidate is the ionized impurity scattering in the well due to the bulk doping. The order of magnitude of this impurity scattering can actually be roughly estimated numerically using the equations in [105]: for this sample, bulk doped at 5×10^{10} cm^{-2}, the calculations show an impurity broadening in the range of 5 % or more. This should not be taken as more than an order-of-magnitude estimate, but it shows that the impurity scattering could plausibly be the reason for the poor linewidth we observe.

9.3 Shifted half parabolic wells

At this point, it might not look promising for HPQWs. Modulation doping leads to a band edge distortion which degrades the absorption quality. Bulk doping does not have such problems, but moving the Si ions into the well likely degrades the absorption quality through ionized impurity scattering. One may wonder, however, if there is a way to combine the strengths of these two methods. That is, can we compensate for the Hartree potential created by the electrons without introducing scatterers into the well region?

To do this, we have designed modulation doped HPQWs which include a slight shift of the half-parabolic profile, so that it is not truncated exactly at the parabola minimum. The inspiration for this idea comes from thinking more carefully about what the bulk doping actually achieves. When we bulk dope a HPQW, we are adding a uniform positive charge density to the well. From Poisson's equation, we know that a uniform charge density will create a parabolic potential profile (see footnote 6, page 51). This will be added

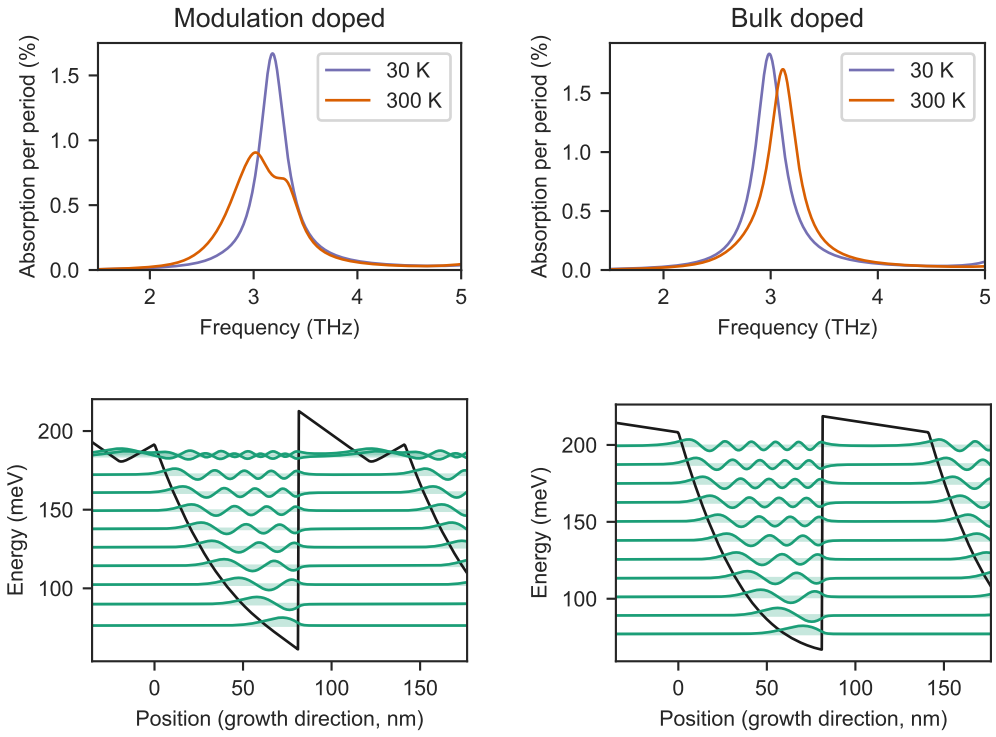

Figure 9.2: Absorption and wavefunctions for the same HPQW system as Fig. 9.1, but doped at 1×10^{11} **Left panels**: the HPQWs are modulation doped with a Si delta part way through the barrier. **Right panels**: the HPQWs are bulk doped uniformly throughout the graded QW region. **Top panels**: absorption calculated using the semiclassical method of Ch. 5 at 30 K and 300 K. For the sake of comparison, a constant linewidth $\gamma = 0.3$ THz is used – i.e., any broadening observed is an inhomogeneous broadening due to the presence of several absorption modes. **Bottom panels**: Wavefunctions and conduction band edges at 300 K, calculated self-consistently using the Schrödinger-Poisson method. Energy is measured relative to the Fermi level, and wavefunctions are offset according to their corresponding subband minima.

on to the bare well potential, which is also parabolic in the absence of doping (although it is truncated so that half the parabola is missing). So, at least within the well region, the overall potential seen by the electrons is that of two parabolas added together. Furthermore, we know that mathematically adding two parabolas together yields a third parabola – just with a different centre and curvature. Thus, we can expect that the effect of placing positive ions throughout the well is to deepen the well and also shift the parabolic potential in the $-z$ direction.

We do not necessarily want to mimic this effect exactly, but we can try to achieve a similar modification of the conduction band in modulation doped wells by altering the composition profile. Rather than using a perfect half-parabola, we use a shifted parabolic composition profile: some examples are shown in Fig. 9.3 (left panel). Mathematically, we can define a family of shifted half-parabolic composition profiles by

$$x(z) = (x_{\max} - x_{\min})\left(\frac{z - z_s}{L_{\mathrm{QW}} - z_s}\right)^2 + x_{\min}; \quad 0 < z < L_{\mathrm{QW}} \tag{9.1}$$

Here x is the $\mathrm{Al}_x\,\mathrm{Ga}_{1-x}\,\mathrm{As}$ composition along the growth direction, z. L is the length of the QW region, and z_s is the shift that we have applied. The minimum of the parabola is located at $z = z_s$. The profile begins at $z = 0$, and is truncated at $z = L_{\mathrm{QW}}$ (above which, we use a constant-composition barrier with $x = x_{\max}$). We have chosen the curvature to be dependent on z_s: this choice ensures that the composition always ranges from $x_{\min}$ to $x_{\max}$ for any choice of z_s applied. Also note that, as we vary the shift z_s from 0 to $L_{\mathrm{QW}}/2$, the structure morphs from a perfect HPQW into a perfect PQW.

By adding a shift to the well bottom, we can offset the potential created by the electrons in the well. An example of this is shown in Fig. 9.3 for a ∼3.0 THz HPQW doped at $1 \times 10^{11}\,\mathrm{cm}^{-2}$. When there is no shift applied (an ideal HPQW), we can see that the conduction band profile ends up being almost triangular due to the Poisson potential of the electrons. If we shift the well bottom by $z_s \sim 10\,\mathrm{nm}$, then we can see that the Poisson potential of the electrons is more or less compensated – the conduction band edge *after* including the electron Poisson potential is much more like an ideal HPQW. As the shift moves higher, it appears that we overcompensate, and a PQW-like minimum begins to appear.

This intuition from the conduction band behaviour is borne out in the absorption profiles at 300 K as well: see Fig. 9.4. With a shift of ∼10 nm we see a very clean absorption peak. This small shift creates a dramatic improvement in the absorption behaviour, and we can see that the intensity of this absorption at 300 K compares favourably to the 30 K absorption of Fig. 9.2. If we use larger shifts, we see that the absorption behaviour remains quite good, although it drops off slightly as we overcompensate for the electron Poisson potential. Of course, if we push the shift even further, at some point we will effectively just return to the case of a full PQW.

The optimal shift will depend on the width of the well and the amount of doping. In general, we should use the full modelling shown here to find the optimum doping level, taking into account both the absorption behaviour and the conduction band behaviour. However, we can come up with an order-of-magnitude guess as a starting point. We know that the maximum doping in a PQW occurs approximately when its plasma frequency (recall: $\omega_P^2 = e^2 N_{2\mathrm{D}}/L_{\mathrm{QW}} m^* \epsilon_0 \epsilon_s$) is equal to its bare transition frequency, ω_0. We can assume that at this maximum doping level, it is necessary to use a full PQW: thus the optimal shift should be $z_s = L_{\mathrm{QW}}/2$ when $\omega_P = \omega_0$. Conversely, if there is no doping, we do not need to apply a shift at all. Using the simplest possible formula to interpolate between these two extremes, we can estimate the optimum shift as

$$z_{s,\mathrm{opt}} \sim \frac{\omega_P}{\omega_0} \cdot \frac{L_{\mathrm{QW}}}{2} \tag{9.2}$$

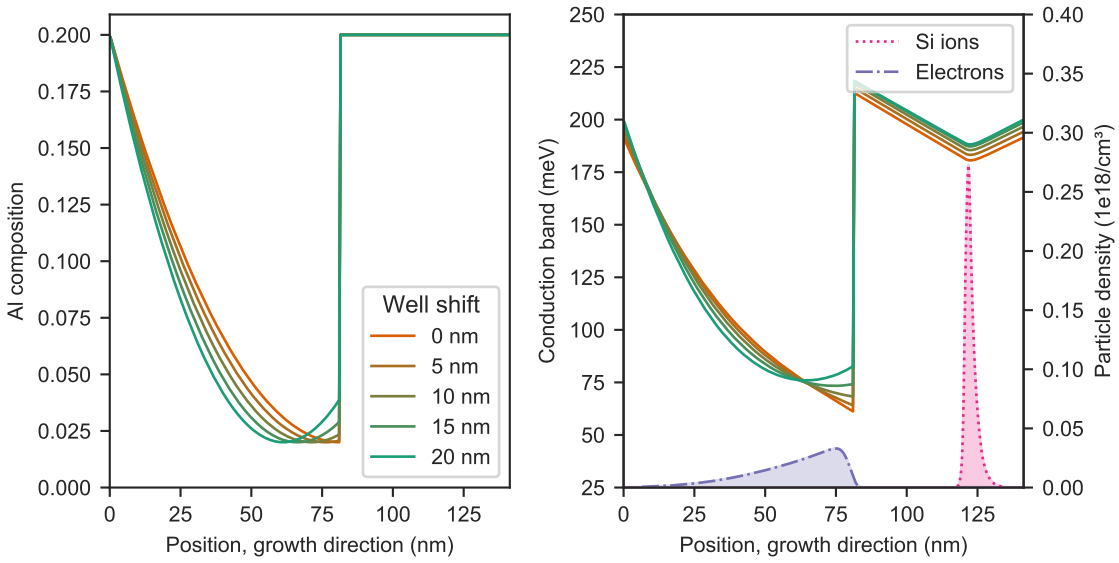

Figure 9.3: Left: Shifted HPQW composition profiles generated by Eq. (9.1). As with the structure of Fig. 9.1, we use a well width of 81.28 nm. **Right**: The calculated conduction band edges (at 300 K) for the shifted profiles, with a modulation doping of 1×10^{11} cm^{-2} per well in the barrier. For reference, the Schrödinger-Poisson calculated electron density (blue) and Si ion density (pink) are shown for the case of $z_s = 0$. The Si density corresponds to a delta doping profile which has been asymmetrically smeared to account for Si segregation during molecular beam epitaxy (MBE) growth [93].

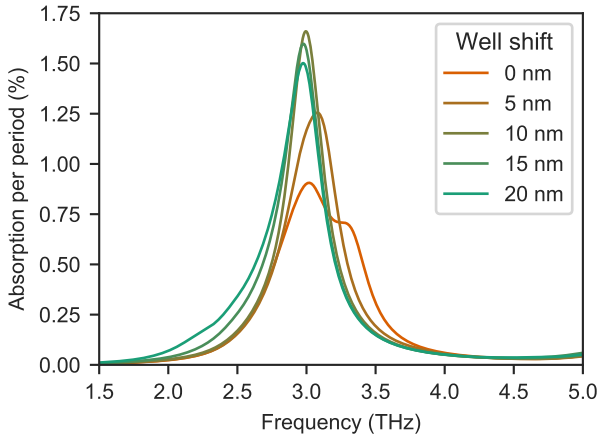

Figure 9.4: Absorption spectra for the shifted HPQWs shown in Fig. 9.3. Again, these wells are modulation doped at 1×10^{11} cm^{-2} per well. Absorption is calculated using the semiclassical method of Ch. 5 at 300 K. For the sake of comparison, a constant linewidth $\gamma = 0.3$ THz is used.

Even though the justification is far from rigorous, empirically, this rule appears to give a reasonable estimate in many cases. For the HPQW shown in Figs. 9.3 and 9.4, for example, this formula gives $z_{s,\mathrm{opt}} \sim 12.6\,\mathrm{nm}$ while the actual optimal shift is around 11.1 nm.

9.4 A test growth

To test this strategy of shifting the well centre, we grew a periodic array of 43 $Al_x Ga_{1-x} As$ HPQWs, modulation doped at $5 \times 10^{10}\,\mathrm{cm}^{-2}$, with a shift of $z_s = 5.7\,\mathrm{nm}$. (This was found to be the optimal shift for this doping level.) As in the simulations above, the wells had composition ranging from $x = 0.02$ to $x = 0.20$, and they were 81.28 nm wide, with 60 nm $Al_{0.2} Ga_{0.8} As$ barriers between them. The wells were grown as a continuous alloy using the technique of Ch. 7.

We used the same multipass absorption setup as in Ch. 8 to measure the absorption. The results are shown in Fig. 9.5. The absorption behaviour is clearly improved compared to the bulk-doped sample (which had a linewidth around 18 %) at 77 K. At cryogenic temperatures we see linewidths better than 10 %. Unfortunately, the linewidths are still not as good as the full PQW samples, and experiments above 150 K may not be feasible with this particular sample. However, it is difficult to draw firm conclusions from this sample alone. On the MBE side, beam flux-monitoring ion gauge (BFM) measurements before and after the growth showed serious cell flux variations of several percent. It is difficult to know when exactly these instabilities occurred during the growth, but it is certainly possible that part of the observed linewidth is an inhomogeneous broadening due to cell instabilities during the QW superlattice growth. Still, these are promising initial results, and further study is underway to determine how viable these shifted HPQWs are for polariton scattering.

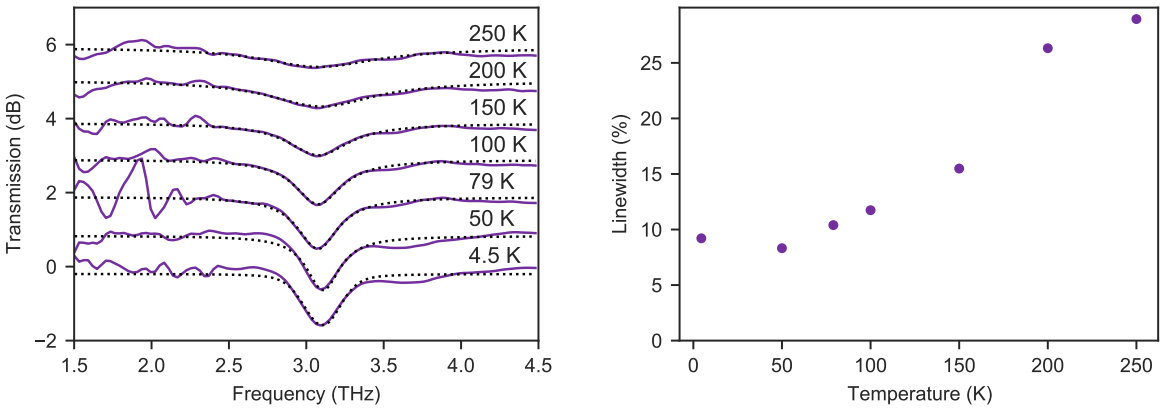

Figure 9.5: Left: multipass transmission of the HPQW sample (growth number G0712 in the QNC-MBE database) at various temperatures. The 300 K measurement is excluded because it was too weak to see clearly in the cryostat setup. (These measurements were performed in the older cryostat setup of Fig. 8.3, which is why the noise is larger than in Fig. 8.8.) The solid purple lines show the measured curves, and the dotted black lines show the best-fit to Eq. 8.1, using the same transfer matrix-style method as in Ch. 8. For clarity, the curves have been offset by 1 dB. **Right**: linewidths, γ, extracted from measurements by the fitting procedure, expressed as a fraction of the centre frequency (3.1 THz).

Chapter 10

Conclusions

In this work, we have examined the development of quantum well (QW) active regions for terahertz (THz) polaritonic devices. Theoretically, we have explored several types of active region design with a numerical model that accounts for the formation of multisubband plasmon (MSP) modes. Experimentally, we have developed a new technique for growing continuously graded alloys in molecular beam epitaxy (MBE), which we used to grow parabolic quantum well (PQW) arrays with high-quality absorption peaks in the THz.

In Chs. 4 and 5, we built up the tools required to model and design THz polariton devices, with a particular emphasis on the QW active regions. We saw that polariton scattering is an important problem of interest, and we saw that high-quality QW absorption resonances are a key component of pursuing such studies and – eventually – for pursuing practical devices. We then examined theoretical models of QW absorption. By combining a standard Schrödinger-Poisson modeller, a small $\mathbf{k} \cdot \mathbf{p}$ correction to the energy dispersions, and a semiclassical plasmon model from the literature, we were able to create a full numerical model for simulating these active regions. We discussed the limitations of this model and the need for future work, and we explored the differences between square and parabolic QWs for THz polaritons. We saw why square wells are problematic at room temperature, and why PQWs should be more promising.

In Chs. 6 and 7, we addressed the growth of PQWs in $\mathrm{Al}_x\,\mathrm{Ga}_{1-x}\,\mathrm{As}$. We developed new growth techniques which allow the alloy composition to be continuously varied during growth in a precise, controllable manner. This allows us to avoid the use of digital alloys, which come with a large number of interfaces. Further, unlike previous work in continuous alloys, our growth technique allows for precise composition control at the high growth rates required for thick structures with many QW repeats. The technique is quite general, and could be applied to other cells and materials.

In Ch. 8, we used this continuous alloy technique to grow high-quality PQWs. We measured their multipass absorption and found that they had remarkably small linewidths, as low as 3.9 %. We saw that the absorption line was robust, and remained relatively narrow even up to 150 K. We also saw evidence of strong coupling within dispersive metal-insulator-metal (MIM) cavities.

Finally, in Ch. 9, we looked forward towards an alternative active region design comprising half PQWs. Using our numerical model, we saw that absorption in these wells is not as naturally robust as in their full PQW cousins. However, we showed theoretically that a small correction to the composition profile should largely mitigate this problem, potentially giving us another choice for high quality THz resonances above cryogenic temperatures. Although we had difficulties with an unstable MBE growth, we grew preliminary

samples which showed some promise.

The primary contributions of this work are fourfold. First is the development of numerical code for QW absorption modelling. While most of the theoretical discussion in Ch. 5 had already existed in the literature, it was critical to combine the pieces and develop code to perform calculations, so that devices could be designed and tested. This work also laid the foundation for further theoretical work in extending the model to include $\mathbf{k} \cdot \mathbf{p}$ non-parabolicity – work that is already underway.

The second main contribution is the continuous alloy growth technique. This technique allows for unparalleled precision at standard growth rates, and it is not limited to PQWs for THz polaritons. It could be applied to other composition profiles, other material systems, or other situations where precise time-varying flux control is needed. The technique does not require special mechanical or electrical equipment, so could be useful to the broader MBE community.

The third contribution is our continuously graded THz PQW active regions. These structures exhibit remarkably low linewidths at low temperature, and this quality is largely maintained even up to 150 K or so. Practically, one result of this is that experiments which previously required liquid helium temperatures (4 K) with square wells could now be comfortably performed with liquid nitrogen (78 K), which is much cheaper and more readily accessible. With a small amount of improvement in the growth and design, these experiments could potentially be performed up to room temperature.

The final contribution is the idea of using shifted half parabolic quantum wells (HPQWs) for THz active regions. HPQWs are not completely unstudied in the literature, but have not been explored in this context. The shifted version we propose could be used in polariton experiments which require heavy doping. They could be an interesting alternative to PQWs, especially in the context of non-linear optical studies.

References

[1]M. Tonouchi, "Cutting-edge terahertz technology", Nature Photonics **1**, 97–105 (2007).

[2]Y.-S. Lee, *Principles of terahertz science and technology*, Vol. 170 (Springer Science & Business Media, 2009).

[3]D. M. Mittleman, "Perspective: terahertz science and technology", Journal of Applied Physics **122**, 230901 (2017).

[4]S. S. Dhillon, M. S. Vitiello, E. H. Linfield, A. G. Davies, C. H. Matthias, B. John, P. Claudio, M. Gensch, P. Weightman, G. P. Williams, E. Castro-Camus, D. R. S. Cumming, F. Simoens, I. Escorcia-Carranza, J. Grant, L. Stepan, K.-G. Makoto, K. Kuniaki, K. Martin, A. S. Charles, L. C. Tyler, H. Rupert, A. G. Markelz, Z. D. Taylor, P. W. Vincent, J. A. Zeitler, S. Juraj, M. K. Timothy, B. Ellison, S. Rea, P. Goldsmith, B. C. Ken, A. Roger, D. Pardo, P. G. Huggard, V. Krozer, S. Haymen, F. Martyn, R. Cyril, S. Alwyn, S. Andreas, N. Mira, R. Nick, C. Roland, E. C. John, and B. J. Michael, "The 2017 terahertz science and technology roadmap", Journal of Physics D: Applied Physics **50**, 043001 (2017).

[5]T. Nagatsuma, G. Ducournau, and C. C. Renaud, "Advances in terahertz communications accelerated by photonics", Nature Photonics **10**, 371–379 (2016).

[6]M. D. Eisaman, J. M. A. P. S. Fan, A. Migdall, and S. V. Polyakov, "Invited review article: single-photon sources and detectors", Review of Scientific Instruments **82**, 071101–071101 (2011).

[7]G. S. Buller and R. J. Collins, "Single-photon generation and detection", Measurement Science and Technology **21**, 012002–012002 (2009).

[8]N. Somaschi, V. Giesz, L. De Santis, J. Loredo, M. P. Almeida, G. Hornecker, S. L. Portalupi, T. Grange, C. Antón, and J. Demory, "Near-optimal single-photon sources in the solid state", Nature Photonics **10**, 340 (2016).

[9]C. Ciuti, G. Bastard, and I. Carusotto, "Quantum vacuum properties of the intersubband cavity polariton field", Physical Review B **72**, 115303–115303 (2005).

[10]P. Dean, A. Valavanis, J. Keeley, K. Bertling, Y. L. Lim, R. Alhathlool, A. D. Burnett, L. H. Li, S. P. Khanna, D. Indjin, et al., "Terahertz imaging using quantum cascade lasers—a review of systems and applications", Journal of Physics D: Applied Physics **47**, 374008–374008 (2014).

[11]T. L. Cocker, V. Jelic, M. Gupta, S. J. Molesky, J. A. J. Burgess, G. D. L. Reyes, L. V. Titova, Y. Y. Tsui, M. R. Freeman, and F. A. Hegmann, "An ultrafast terahertz scanning tunnelling microscope", Nature Photonics **7**, 620–625 (2013).

[12]D. Saeedkia, *Handbook of terahertz technology for imaging, sensing and communications* (Elsevier, 2013).

[13]J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, A. Y. Cho, et al., "Quantum cascade laser", Science **264**, 553–556 (1994).

[14]R. Köhler, A. Tredicucci, F. Beltram, H. E. Beere, E. H. Linfield, A. G. Davies, D. A. Ritchie, R. C. Iotti, and F. Rossi, "Terahertz semiconductor-heterostructure laser", Nature **417**, 156–159 (2002).

[15]A. Khalatpour, A. K. Paulsen, C. Deimert, Z. R. Wasilewski, and Q. Hu, "High-power portable terahertz laser systems", Nature Photonics **15**, 16–20 (2021).

[16]Y. Chassagneux, Q. J. Wang, S. P. Khanna, E. Strupiechonski, J. Coudevylle, E. H. Linfield, A. G. Davies, F. Capasso, M. A. Belkin, and R. Colombelli, "Limiting factors to the temperature performance of THz quantum cascade lasers based on the resonant-phonon depopulation scheme", IEEE Transactions on Terahertz Science and Technology **2**, 83–92 (2012).

[17]S. Fathololoumi, E. Dupont, C. W. I. Chan, Z. R. Wasilewski, S. R. Laframboise, D. Ban, A. Mátyás, C. Jirauschek, Q. Hu, and H. C. Liu, "Terahertz quantum cascade lasers operating up to 200 K with optimized oscillator strength and improved injection tunneling", Optics Express **20**, 3866–3876 (2012).

[18]L. Bosco, M. Franckié, G. Scalari, M. Beck, A. Wacker, and J. Faist, "Thermoelectrically cooled THz quantum cascade laser operating up to 210 K", Applied Physics Letters **115**, 010601 (2019).

[19]M. A. Belkin and F. Capasso, "New frontiers in quantum cascade lasers: high performance room temperature terahertz sources", Physica Scripta **90**, 118002–118002 (2015).

[20]M. Wienold, B. Röben, L. Schrottke, R. Sharma, A. Tahraoui, K. Biermann, and H. T. Grahn, "High-temperature, continuous-wave operation of terahertz quantum-cascade lasers with metal-metal waveguides and third-order distributed feedback", Optics Express **22**, 3334–3348 (2014).

[21]B. S. Williams, "Terahertz quantum-cascade lasers", Nature Photonics **1**, 517–525 (2007).

[22]C. Sirtori, S. Barbieri, and R. Colombelli, "Wave engineering with THz quantum cascade lasers", Nature Photonics **7**, 691–701 (2013).

[23]A. Kavokin, J. J. Baumberg, G. Malpuech, and F. P. Laussy, *Microcavities* (Oxford University Press, 2007).

[24]C. Weisbuch, M. Nishioka, A. Ishikawa, and Y. Arakawa, "Observation of the coupled exciton-photon mode splitting in a semiconductor quantum microcavity", Physical Review Letters **69**, 3314–3317 (1992).

[25]D. Bajoni, "Polariton lasers. hybrid light–matter lasers without inversion", Journal of Physics D: Applied Physics **45**, 313001–313001 (2012).

[26]H. Haug and S. W. Koch, *Quantum theory of the optical and electronic properties of semiconductors* (World Scientific, 2009).

[27]Y. Yamamoto and A. Imamoglu, *Mesoscopic quantum optics* (John Wiley & Sons, Inc., 1999).

[28]H. Deng, H. Haug, and Y. Yamamoto, "Exciton-polariton Bose-Einstein condensation", Reviews of Modern Physics **82**, 1489–1489 (2010).

[29]B. Deveaud, "Exciton-polariton Bose-Einstein condensates", Annual Review of Condensed Matter Physics **6**, 155–175 (2015).

[30]T. Byrnes, N. Y. Kim, and Y. Yamamoto, "Exciton-polariton condensates", Nature Physics **10**, 803–813 (2014).

[31]I. Carusotto and C. Ciuti, "Quantum fluids of light", Reviews of Modern Physics **85**, 299–299 (2013).

[32]T. C. H. Liew, I. A. Shelykh, and G. Malpuech, "Polaritonic devices", Physica E: Low-dimensional Systems and Nanostructures **43**, 1543–1568 (2011).

[33]M. Combescot, O. Betbeder-Matibet, and F. Dubin, "The many-body physics of composite bosons", Physics Reports **463**, 215–320 (2008).

[34]S. Christopoulos, G. B. H. Von Högersthal, A. J. D. Grundy, P. G. Lagoudakis, A. V. Kavokin, J. J. Baumberg, G. Christmann, R. Butté, E. Feltin, and J. F. Carlin, "Room-temperature polariton lasing in semiconductor microcavities", Physical Review Letters **98**, 126405–126405 (2007).

[35]M. H. Anderson, J. R. Ensher, M. R. Matthews, C. E. Wieman, and E. A. Cornell, "Observation of Bose-Einstein condensation in a dilute atomic vapor", Science **269**, 198–201 (1995).

[36]K. B. Davis, M. O. Mewes, M. R. Andrews, N. J. Van Druten, D. S. Durfee, D. M. Kurn, and W. Ketterle, "Bose-Einstein condensation in a gas of sodium atoms", Physical Review Letters **75**, 3969–3969 (1995).

[37]J. Klaers, J. Schmitt, F. Vewinger, and M. Weitz, "Bose-Einstein condensation of photons in an optical microcavity", Nature **468**, 545–545 (2010).

[38]J. Kasprzak, M. Richard, S. Kundermann, A. Baas, P. Jeambrun, J. M. J. Keeling, F. M. Marchetti, M. H. Szymańska, R. Andre, and J. L. Staehli, "Bose-Einstein condensation of exciton polaritons", Nature **443**, 409–414 (2006).

[39]P. G. Savvidis, J. J. Baumberg, R. M. Stevenson, M. S. Skolnick, D. M. Whittaker, and J. S. Roberts, "Angle-resonant stimulated polariton amplifier", Physical Review Letters **84**, 1547–1547 (2000).

[40]A. Imamoglu, R. J. Ram, S. Pau, and Y. Yamamoto, "Nonequilibrium condensates and lasers without inversion: exciton-polariton lasers", Physical Review A **53**, 4250–4250 (1996).

[41]C. Schneider, A. Rahimi-Iman, N. Y. Kim, J. Fischer, I. G. Savenko, M. Amthor, M. Lermer, A. Wolf, L. Worschech, V. D. Kulakovskii, et al., "An electrically pumped polariton laser", Nature **497**, 348–352 (2013).

[42]P. Bhattacharya, B. Xiao, A. Das, S. Bhowmick, and J. Heo, "Solid state electrically injected exciton-polariton laser", Physical Review Letters **110**, 206403–206403 (2013).

[43]P. Bhattacharya, T. Frost, S. Deshpande, M. Z. Baten, A. Hazari, and A. Das, "Room temperature electrically injected polariton laser", Physical Review Letters **112**, 236802–236802 (2014).

[44]D. Sanvitto and S. Kéna-Cohen, "The road towards polaritonic devices", Nature Materials **15**, 1061–1073 (2016).

[45]S. Kim, B. Zhang, Z. Wang, J. Fischer, S. Brodbeck, M. Kamp, C. Schneider, S. Höfling, and H. Deng, "Coherent polariton laser", Physical Review X **6**, 011026 (2016).

[46]K. V. Kavokin, M. A. Kaliteevski, R. A. Abram, A. V. Kavokin, S. Sharkova, and I. A. Shelykh, "Stimulated emission of terahertz radiation by exciton-polariton lasers", Applied Physics Letters **97**, 201111–201111 (2010).

[47]A. V. Kavokin, I. A. Shelykh, T. Taylor, and M. M. Glazov, "Vertical cavity surface emitting terahertz laser", Physical Review Letters **108**, 197401–197401 (2012).

[48]K. Kristinsson, O. Kyriienko, T. Liew, and I. Shelykh, "Continuous terahertz emission from dipolaritons", Physical Review B **88**, 245303 (2013).

[49]O. Kyriienko, A. Kavokin, and I. Shelykh, "Superradiant terahertz emission by dipolaritons", Physical Review Letters **111**, 176401 (2013).

[50]K. Kristinsson, O. Kyriienko, and I. Shelykh, "Terahertz laser based on dipolaritons", Physical Review A **89**, 023836 (2014).

[51]M. A. Kaliteevskii and K. A. Ivanov, "Double-boson stimulated terahertz emission in a polariton cascade laser", Technical Physics Letters **39**, 91–94 (2013).

[52]T. C. H. Liew, M. Glazov, K. Kavokin, I. A. Shelykh, M. Kaliteevski, and A. V. Kavokin, "Proposal for a bosonic cascade laser", Physical Review Letters **110**, 047402 (2013).

[53]A. Tzimis, A. V. Trifonov, G. Christmann, S. I. Tsintzos, Z. Hatzopoulos, I. V. Ignatiev, A. V. Kavokin, and P. G. Savvidis, "Strong coupling and stimulated emission in single parabolic quantum well microcavity for terahertz cascade", Applied Physics Letters **107**, 101101–101101 (2015).

[54]T. C. H. Liew, Y. G. Rubo, A. S. Sheremet, S. De Liberato, I. A. Shelykh, F. P. Laussy, and A. V. Kavokin, "Quantum statistics of bosonic cascades", New Journal of Physics **18**, 023041–023041 (2016).

[55]S. Huppert, O. Lafont, E. Baudin, J. Tignon, and R. Ferreira, "Terahertz emission from multiple-microcavity exciton-polariton lasers", Physical Review B **90**, 241302 (2014).

[56]B. Askenazi, A. Vasanelli, Y. Todorov, E. Sakat, J.-J. Greffet, G. Beaudoin, I. Sagnes, and C. Sirtori, "Midinfrared ultrastrong light–matter coupling for THz thermal emission", ACS Photonics **4**, 2550–2555 (2017).

[57]S. De Liberato, "Virtual photons in the ground state of a dissipative system", Nature Communications **8**, 1465 (2017).

[58]S. De Liberato and C. Ciuti, "Stimulated scattering and lasing of intersubband cavity polaritons", Physical Review Letters **102**, 136403–136403 (2009).

[59]R. Colombelli and J.-M. Manceau, "Perspectives for intersubband polariton lasers", Physical Review X **5** (2015).

[60]M. Geiser, G. Scalari, F. Castellano, M. Beck, and J. Faist, "Room temperature terahertz polariton emitter", Applied Physics Letters **101**, 141118–141118 (2012).

[61]M. Załużny and C. Nalewajko, "Coupling of infrared radiation to intersubband transitions in multiple quantum wells: the effective-medium approach", Physical Review B **59**, 13043–13053 (1999).

[62]A. A. Anappara, D. Barate, A. Tredicucci, J. Devenson, R. Teissier, and A. Baranov, "Giant intersubband polariton splitting in InAs/AlSb microcavities", Solid State Communications **142**, 311–313 (2007).

[63]A. A. Anappara, S. De Liberato, A. Tredicucci, C. Ciuti, G. Biasiol, L. Sorba, and F. Beltram, "Signatures of the ultrastrong light-matter coupling regime", Physical Review B **79**, 201303–201303 (2009).

[64]P. Jouy, A. Vasanelli, Y. Todorov, L. Sapienza, R. Colombelli, U. Gennser, and C. Sirtori, "Intersubband electroluminescent devices operating in the strong-coupling regime", Physical Review B **82**, 045322 (2010).

[65]S. Zanotto, F. P. Mezzapesa, F. Bianco, G. Biasiol, L. Baldacci, M. S. Vitiello, L. Sorba, R. Colombelli, and A. Tredicucci, "Perfect energy-feeding into strongly coupled systems and interferometric control of polariton absorption", Nature Physics **10**, 830–834 (2014).

[66]M. Geiser, G. Scalari, M. Beck, C. Walther, and J. Faist, "Terahertz LC microcavities: from quantum cascade lasers to ultrastrong light-matter coupling", Journal of Infrared, Millimeter, and Terahertz Waves **34**, 393–404 (2013).

[67]M. Geiser, C. Walther, G. Scalari, M. Beck, M. Fischer, L. Nevou, and J. Faist, "Strong light-matter coupling at terahertz frequencies at room temperature in electronic LC resonators", Applied Physics Letters **97**, 191107 (2010).

[68]M. Geiser, F. Castellano, G. Scalari, M. Beck, L. Nevou, and J. Faist, "Ultrastrong coupling regime and plasmon polaritons in parabolic semiconductor quantum wells", Physical Review Letters **108**, 106402–106402 (2012).

[69]M. Geiser, M. Beck, and J. Faist, "Terahertz intersubband polariton tuning by electrical gating", Optics Express **22**, 2126–2131 (2014).

[70]A. P. Hibbins, J. R. Sambles, C. R. Lawrence, and J. R. Brown, "Squeezing millimeter waves into microns", Physical Review Letters **92**, 143904 (2004).

[71]Y. Todorov, L. Tosetto, J. Teissier, A. M. Andrews, P. Klang, R. Colombelli, I. Sagnes, G. Strasser, and C. Sirtori, "Optical properties of metal-dielectric-metal microcavities in the THz frequency range", Optics Express **18**, 13886–13907 (2010).

[72]Y. Todorov, A. M. Andrews, R. Colombelli, S. De Liberato, C. Ciuti, P. Klang, G. Strasser, and C. Sirtori, "Ultrastrong light-matter coupling regime with polariton dots", Physical Review Letters **105**, 196402 (2010).

[73]P. Jouy, A. Vasanelli, Y. Todorov, A. Delteil, G. Biasiol, L. Sorba, and C. Sirtori, "Transition from strong to ultrastrong coupling regime in mid-infrared metal-dielectric-metal cavities", Applied Physics Letters **98**, 231114 (2011).

[74]A. Vasanelli, Y. Todorov, and C. Sirtori, "Ultra-strong light–matter coupling and superradiance using dense electron gases", Comptes Rendus Physique **17**, 861–873 (2016).

[75]J. M. Manceau, S. Zanotto, T. Ongarello, L. Sorba, A. Tredicucci, G. Biasiol, and R. Colombelli, "Mid-infrared intersubband polaritons in dispersive metal-insulator-metal resonators", Applied Physics Letters **105**, 081105–081105 (2014).

[76]S. L. Chuang, *Physics of photonic devices*, 2nd (John Wiley & Sons Inc., Hoboken, New Jersey, 2008).

[77]J. M. Manceau, S. Zanotto, I. Sagnes, G. Beaudoin, and R. Colombelli, "Optical critical coupling into highly confining metal-insulator-metal resonators", Applied Physics Letters **103**, 091110–091110 (2013).

[78]Y. Todorov and C. Sirtori, "Intersubband polaritons in the electrical dipole gauge", Physical Review B **85**, 10.1103/PhysRevB.85.045304 (2012).

[79]Y. Todorov, "Dipolar quantum electrodynamics of the two-dimensional electron gas", Physical Review B **91**, 125409–125409 (2015).

[80]P. Lalanne and G. M. Morris, "Highly improved convergence of the coupled-wave method for TM polarization", Journal of the Optical Society of America A **13**, 779–784 (1996).

[81]L. Li, "Use of fourier series in the analysis of discontinuous periodic structures", Journal of the Optical Society of America A **13**, 1870–1876 (1996).

[82]D. M. Whittaker and I. S. Culshaw, "Scattering-matrix treatment of patterned multilayer photonic structures", Physical Review B **60**, 2610–2618 (1999).

[83]S. Zanotto, G. Biasiol, R. Degl'Innocenti, L. Sorba, and A. Tredicucci, "Intersubband polaritons in a one-dimensional surface plasmon photonic crystal", Applied Physics Letters **97**, 231123–231123 (2010).

[84]A. Delteil, A. Vasanelli, P. Jouy, D. Barate, J. C. Moreno, R. Teissier, A. N. Baranov, and C. Sirtori, "Optical phonon scattering of cavity polaritons in an electroluminescent device", Physical Review B **83**, 081404–081404 (2011).

[85]L. Nguyen-thê, S. De Liberato, M. Bamba, and C. Ciuti, "Effective polariton-polariton interactions of cavity-embedded two-dimensional electron gases", Physical Review B **87**, 235322–235322 (2013).

[86]J. Nespolo and I. Carusotto, "Generalized Gross-Pitaevskii model for intersubband polariton lasing", Physical Review B **100**, 035305 (2019).

[87]S. De Liberato and C. Ciuti, "Quantum theory of electron tunneling into intersubband cavity polariton states", Physical Review B **79**, 075317–075317 (2009).

[88]B. Paulillo, J. M. Manceau, L. H. Li, A. G. Davies, E. H. Linfield, and R. Colombelli, "Room temperature strong light-matter coupling in three dimensional terahertz meta-atoms", Applied Physics Letters **108**, 101101–101101 (2016).

[89]P. Goulain, A. D. Koulouklidis, J.-M. Manceau, C. Daskalaki, B. Paulillo, K. Maussang, S. Dhillon, J. R. Freeman, L. Li, E. H. Linfield, S. Tzortzakis, and R. Colombelli, "Femtosecond broadband frequency switch of terahertz three-dimensional meta-atoms", ACS Photonics **8**, 1097–1102 (2021).

[90]M. Helm, "The basic physics of intersubband transitions", in *Intersubband transitions in quantum wells: physics and device applications*, edited by H. Liu and F. Capasso (Academic Press, San Diego, 1999) Chap. 1.

[91]S. Birner, T. Zibold, T. Andlauer, T. Kubis, M. Sabathil, A. Trellakis, and P. Vogl, "Nextnano: general purpose 3-D simulations", IEEE Transactions on Electron Devices **54**, 2137–2142 (2007).

[92]S. Birner, *Modeling of semiconductor nanostructures and semiconductor-electrolyte interfaces* (Verein zur Förderung des Walter-Schottky-Inst. der Techn. Univ. München, 2011).

[93]Z. R. Wasilewski, H. C. Liu, and M. Buchanan, "Studies of Si segregation in GaAs using current-voltage characteristics of quantum well infrared photodetectors", Journal of Vacuum Science & Technology B **12**, 1273–1276 (1994).

[94]R. J. Warburton, C. Gauer, A. Wixforth, J. P. Kotthaus, B. Brar, and H. Kroemer, "Intersubband resonances in InAs/AlSb quantum wells: selection rules, matrix elements, and the depolarization field", Physical Review B **53**, 7903–7910 (1996).

[95]I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, "Band parameters for iii–v compound semiconductors and their alloys", Journal of Applied Physics **89**, 5815–5875 (2001).

[96]T. Ando, A. B. Fowler, and F. Stern, "Electronic properties of two-dimensional systems", Reviews of Modern Physics **54**, 437 (1982).

[97]G. Pegolotti, A. Vasanelli, Y. Todorov, and C. Sirtori, "Quantum model of coupled intersubband plasmons", Physical Review B **90**, 035305 (2014).

[98]Y. Todorov, "Dipolar quantum electrodynamics theory of the three-dimensional electron gas", Physical Review B **89**, 075115 (2014).

[99]Y. Todorov, L. Tosetto, A. Delteil, A. Vasanelli, C. Sirtori, A. M. Andrews, and G. Strasser, "Polaritonic spectroscopy of intersubband transitions", Physical Review B **86**, 125314–125314 (2012).

[100]B. Askenazi, A. Vasanelli, A. Delteil, Y. Todorov, L. C. Andreani, G. Beaudoin, I. Sagnes, and C. Sirtori, "Ultra-strong light–matter coupling for designer Reststrahlen band", New Journal of Physics **16**, 043029 (2014).

[101]F. Alpeggiani and L. C. Andreani, "Semiclassical theory of multisubband plasmons: nonlocal electrodynamics and radiative effects", Physical Review B **90**, 115311 (2014).

[102]L. Wendler and E. Kändler, "Intra- and intersubband plasmon-polaritons in semiconductor quantum wells", physica status solidi (b) **177**, 9–67 (1993).

[103]M. Załużny, "On the optical properties of quasi-two-dimensional electrons", Physica B+C **128**, 171–182 (1985).

[104]F. Tassone, F. Bassani, and L. Andreani, "Resonant and surface polaritons in quantum wells", Il Nuovo Cimento D **12**, 1673–1687 (1990).

[105]T. Unuma, M. Yoshita, T. Noda, H. Sakaki, and H. Akiyama, "Intersubband absorption linewidth in GaAs quantum wells due to scattering by interface roughness, phonons, alloy disorder, and impurities", Journal of Applied Physics **93**, 1586 (2003).

[106]J. M. Manceau, G. Biasiol, N. L. Tran, I. Carusotto, and R. Colombelli, "Immunity of intersubband polaritons to inhomogeneous broadening", Physical Review B **96**, 235301 (2017).

[107]W. Kohn, "Cyclotron resonance and de Haas-van Alphen oscillations of an interacting electron gas", Physical Review **123**, 1242 (1961).

[108]L. Brey, N. F. Johnson, and B. I. Halperin, "Optical and magneto-optical absorption in parabolic quantum wells", Physical Review B **40**, 10647–10649 (1989).

[109]A. Delteil, A. Vasanelli, Y. Todorov, C. F. Palma, M. R. St-Jean, G. Beaudoin, I. Sagnes, and C. Sirtori, "Charge-induced coherence between intersubband plasmons in a quantum structure", Physical Review Letters **109**, 246808 (2012).

[110]M. J. Manfra, "Molecular beam epitaxy of ultra-high-quality AlGaAs/GaAs heterostructures: enabling physics in low-dimensional electronic systems", Annu. Rev. Condens. Matter Phys. **5**, 347–373 (2014).

[111]M. Jeannin, G. Mariotti Nesurini, S. Suffit, D. Gacemi, A. Vasanelli, L. Li, A. G. Davies, E. Linfield, C. Sirtori, and Y. Todorov, "Ultrastrong light–matter coupling in deeply subwavelength thz lc resonators", ACS Photonics **6**, 1207–1215 (2019).

[112]Y. Zhao, J. Zhang, C. Cai, J. Chen, X. Zhao, C. Liang, F. Liu, Y. Shi, X. Liu, and R. Che, "Domino effect of thickness fluctuation on subband structure and electron transport within semiconductor cascade structures", ACS Applied Materials & Interfaces **12**, 41950–41959 (2020).

[113]M. A. Herman and H. Sitter, *Molecular beam epitaxy: fundamentals and current status*, Second, Revised and Updated, Vol. 7 (Springer Science & Business Media, Berlin, 2012).

[114]D. N. Gerasimov and E. I. Yurin, *Kinetics of evaporation*, Vol. 68 (Springer, 2018).

[115]C. Deimert and Z. R. Wasilewski, "MBE growth of continuously-graded parabolic quantum well arrays in AlGaAs", Journal of Crystal Growth **514**, 103–108 (2019).

[116]C. Deimert and Z. R. Wasilewski, "Precise control of time-varying effusion cell flux in molecular beam epitaxy", Journal of Vacuum Science & Technology A **39**, 043407 (2021).

[117]Y. Yao, A. J. Hoffman, and C. F. Gmachl, "Mid-infrared quantum cascade lasers", Nature Photonics **6**, 432–439 (2012).

[118]Z. R. Wasilewski, M. M. Dion, D. J. Lockwood, P. Poole, R. W. Streater, and A. J. SpringThorpe, "Composition of AlGaAs", Journal of Applied Physics **81**, 1683–1694 (1997).

[119]R. C. Miller, A. C. Gossard, D. A. Kleinman, and O. Munteanu, "Parabolic quantum wells with the GaAs-AlGaAs system", Physical Review B **29**, 3740 (1984).

[120]M. Sundaram, A. Wixforth, R. S. Geels, A. C. Gossard, and J. H. English, "A direct method to produce and measure compositional grading in $Al_x Ga_{1-x} As$ alloys", Journal of Vacuum Science & Technology B **9**, 1524–1529 (1991).

[121]M. K. Tucker and D. G. Meyer, "Nonlinear modeling, identification, and feedback control design for the modern effusion cell", Journal of Vacuum Science & Technology A **16**, 3536–3553 (1998).

[122]W. Q. Chen, S. M. Wang, T. G. Andersson, and J. Thordson, "Inverse parabolic quantum wells grown by molecular-beam epitaxy using digital and analog techniques", Physical Review B **48**, 14264–14268 (1993).

[123]S. Wang, G. Treiderис, W. Chen, and T. Andersson, "Growth of analog AlGaAs/GaAs parabolic quantum wells by molecular beam epitaxy", Applied Physics Letters **62**, 61–62 (1993).

[124]S. Giugni and T. Tansley, "Molecular beam epitaxy growth and physical characterization of precise, narrow, triangular heterostructures using an analog grading algorithm", Journal of Vacuum Science & Technology B **9**, 2805–2813 (1991).

[125]D. E. Aspnes, W. E. Quinn, M. C. Tamargo, M. A. A. Pudensi, S. A. Schwarz, M. J. S. P. Brasil, R. E. Nahory, and S. Gregory, "Growth of $Al_xGa_{1-x}As$ parabolic quantum wells by real-time feedback control of composition", Applied Physics Letters **60**, 1244–1246 (1992).

[126]K. R. Evans, R. Kaspi, W. T. Cooley, C. R. Jones, and J. S. Solomon, "Arbitrary composition profiles by MBE using desorption mass spectrometry", MRS Proceedings **300**, 495 (1993).

[127]K. R. Evans, R. Kaspi, C. R. Jones, R. E. Sherriff, V. Jogai, and D. C. Reynolds, "Desorption mass spectrometric control of composition during MBE growth of AlGaAs", Journal of Crystal Growth **127**, 523–527 (1993).

[128]J. Harbison, L. Peterson, and J. Levkoff, "Precisely controlled compositional gradients in MBE grown AlGaAs/GaAs structures", Journal of Crystal Growth **81**, 34–37 (1987).

[129]M. Sundaram, A. Gossard, J. English, and R. Westervelt, "Remotely-doped graded potential well structures", Superlattices and Microstructures **4**, 683–691 (1988).

[130]C. Deimert, P. Goulain, J. M. Manceau, W. Pasek, T. Yoon, A. Bousseksou, N. Y. Kim, R. Colombelli, and Z. R. Wasilewski, "Realization of harmonic oscillator arrays with graded semiconductor quantum wells", Physical Review Letters **125**, 097403 (2020).

[131]W. G. Breiland and K. P. Killeen, "A virtual interface method for extracting growth rates and high temperature optical constants from thin semiconductor films using insitu normal incidence reflectance", Journal of Applied Physics **78**, 6726–6736 (1995).

[132]G. Bastard, E. Mendez, L. Chang, and L. Esaki, "Exciton binding energy in quantum wells", Physical Review B **26**, 1974 (1982).

[133]Q. Zhang, T. Arikawa, E. Kato, J. L. Reno, W. Pan, J. D. Watson, M. J. Manfra, M. A. Zudov, M. Tokman, M. Erukhimova, A. Belyanin, and J. Kono, "Superradiant decay of cyclotron resonance of two-dimensional electron gases", Physical Review Letters **113**, 047601 (2014).

[134]T. Laurent, Y. Todorov, A. Vasanelli, A. Delteil, C. Sirtori, I. Sagnes, and G. Beaudoin, "Superradiant emission from a collective excitation in a semiconductor", Physical Review Letters **115**, 187402 (2015).

[135]R. Harrington, *Time-harmonic electromagnetic fields* (Wiley-IEEE press, New York, NY, 1961).

[136]J. Mathews and R. Howell, *Complex analysis for mathematics and engineering*, 6th (Jones & Bartlett Publishers, 2012).

[137]B. Wen, C. Deimert, S. Wang, C. Xu, S. S. Rassel, Z. Wasilewski, and D. Ban, "Six-level hybrid extraction/injection scheme terahertz quantum cascade laser with suppressed thermally activated carrier leakage", Optics Express **28**, 26499–26508 (2020).

[138]C. R. H. McRae, J. H. Béjanin, C. T. Earnest, T. G. McConkey, J. R. Rinehart, C. Deimert, J. P. Thomas, Z. R. Wasilewski, and M. Mariantoni, "Thin film metrology and microwave loss characterization of indium and aluminum/indium superconducting planar resonators", Journal of Applied Physics **123**, 205304 (2018).

APPENDICES

Appendix A

Electromagnetic derivations for the semiclassical model

In this appendix, we will derive the solutions to the electromagnetic field equations for the semiclassical plasmon model discussed in Sec. 5.4. This section fills in some gaps of the derivation in Ref. [101], and also generalizes their results to quantum wells (QWs) that might not have inversion symmetry. Finally, we will derive an expression for a *local* effective permittivity tensor in the long-wavelength approximation.

A.1 Electromagnetic fields in the quantum well region

We begin with Maxwell's equations, assuming no sources and no magnetic materials:

$$\boldsymbol{\nabla} \times \mathbf{E} = -\mu_0 \partial_t \mathbf{H} \tag{A.1}$$

$$\boldsymbol{\nabla} \times \mathbf{H} = \partial_t \mathbf{D} \tag{A.2}$$

$$\boldsymbol{\nabla} \cdot \mathbf{D} = 0 \tag{A.3}$$

$$\boldsymbol{\nabla} \cdot \mathbf{B} = 0 \tag{A.4}$$

$\mathbf{E}, \mathbf{D}, \mathbf{H}, \mathbf{B}$ are the electric field, electric displacement field, magnetic field, and magnetic flux density, respectively. In general, we would like to solve full structures such as the metal-insulator-metal (MIM) cavities described in Ch. 4, however we will focus here on the QW region, which includes a non-local susceptibility tensor in the growth direction (z). Once we have an effective permittivity for this region, the other regions can be solved using a standard electromagnetic solver.

We assume that the QW region is time-invariant and spatially invariant in the $x-y$ plane. This will allow us to replace $\partial_t \to i\omega, \partial_x \to iq_x, \partial_y \to iq_y$ in the following, where q_x and q_y are the spatial wave numbers. (This is equivalent to taking Fourier transforms of Maxwell's equations.) For brevity of notation, we will continue to use the same field symbols, but we will assume from this point on that all fields have been Fourier transformed along these coordinates and are functions of ω, q_x, q_y. So, for example, $D_z(z) \equiv D_z(\omega, q_x, q_y, z) \equiv \mathscr{F}_{x,y,t}\{D_z(x, y, z, t)\}$.

Further, we can always split the field into two types of modes [135]:

- Transverse magnetic (TM) modes where $H_z = 0$
- Transverse electric (TE) modes where $E_z = 0$

The TE modes are not of interest here, since they will not couple to the quantum well and can be solved using standard techniques. So, we will restrict ourselves to TM modes.

For a TM mode, since $H_z = 0$ by definition, we can write the magnetic field as

$$\mathbf{H} = H_y(z)\hat{\mathbf{y}} \tag{A.5}$$

In principle, the TM mode could have both x and y components, but, since we have assumed invariance in the $x-y$ plane, we can always rotate our coordinate system so that $H_x = 0$.

Now, Eq. A.4 with our TM expression for **H** gives (assuming material with a constant magnetic permeability)

$$q_y H_y(z) = 0 \Longrightarrow q_y = 0 \tag{A.6}$$

So we can conclude that there is no spatial variation in the y direction. (In other words, rotating our coordinate system so that $H_x = 0$ is equivalent to rotating our coordinate system so that the wave is travelling in the $x-z$ plane.)

Plugging in our TM expression for **H**, we can write Eq. A.1 as

$$iq_x E_y = 0 \tag{A.7}$$

$$-\partial_z E_y = 0 \tag{A.8}$$

$$\partial_z E_x - iq_x E_z = -i\omega\mu_0 H_y \tag{A.9}$$

Plugging in our TM expression for **H**, we can write Eq. A.2 as

$$iq_x H_y = i\omega D_z \tag{A.10}$$

$$-\partial_z H_y = i\omega D_x \tag{A.11}$$

We are particularly interested in solving for the z component of the electric field, which is what couples to the QWs. We can rearrange (A.9), (A.10), and (A.11) to obtain

$$q_x \partial_z E_x - iq_x^2 E_z = -i\omega^2 \mu_0 D_z \tag{A.12}$$

$$q_x D_x = i\partial_z D_z \tag{A.13}$$

To solve these, we need a relationship between **D** and **E**. As discussed in Sec. 5.4.1, we use a non-local, anisotropic susceptibility so that

$$D_x = \epsilon_0 \epsilon_{xx} E_x \tag{A.14}$$

$$D_z = \epsilon_0 \epsilon_s E_z + \epsilon_0 \int \chi_{zz}(z, z') E_z(z')\, \mathrm{d}z' \tag{A.15}$$

ϵ_{xx} is the in-plane permittivity and ϵ_s is the background permittivity of the undoped semiconductor. In the simplest case, we would have $\epsilon_{xx} = \epsilon_s =$ constant. However, in general, they could be different from

each other and dependent on ω to account for in-plane plasma oscillations [83] and/or the frequency dependent permittivity of the background semiconductor (e.g., from optical phonons).

In the z direction, let us simplify the notation for the time being by writing

$$D_z = \epsilon_0(\epsilon_s + \chi_{zz})E_z \tag{A.16}$$

with the understanding that χ_{zz} is acting as a linear operator. Plugging these in to (A.12) and simplifying, (assuming $\partial_z \epsilon_{xx} = 0$) we get

$$q_x(\epsilon_s + \chi_{zz})\partial_z D_x - iq_x^2 \epsilon_{xx} D_z = -i\omega^2 \mu_0 \epsilon_0 \epsilon_{xx}(\epsilon_s + \chi_{zz})D_z \tag{A.17}$$

Then, using (A.13), we get

$$(\epsilon_s + \chi_{zz})\partial_z^2 D_z - q_x^2 \epsilon_{xx} D_z = -k_0^2 \epsilon_{xx}(\epsilon_s + \chi_{zz})D_z \tag{A.18}$$

where

$$k_0^2 = \omega^2 \mu_0 \epsilon_0 = \frac{\omega^2}{c^2} \tag{A.19}$$

Rearranging, we get

$$\left(\epsilon_s \partial_z^2 + \epsilon_{xx}\epsilon_s k_0^2 - \epsilon_{xx} q_x^2\right)D_z = -\chi_{zz}\left(\partial_z^2 + \epsilon_{xx}k_0^2\right)D_z \tag{A.20}$$

Defining $k_z^2 = \epsilon_{xx}k_0^2 - (\epsilon_{xx}/\epsilon_s)q_x^2$, and writing out the χ_{zz} operator explicitly, we get the final result

$$\boxed{\left(\partial_z^2 + k_z^2\right)D_z = -\int \frac{\chi_{zz}(z,z')}{\epsilon_s}\left(\partial_{z'}^2 + \epsilon_{xx}k_0^2\right)D_z(z')\,\mathrm{d}z'} \tag{A.21}$$

(Note: our definition of k_z reduces to that of Ref. [101] when $\epsilon_{xx} = \epsilon_s$. Care must be taken when applying boundary conditions in the case $\epsilon_{xx} \neq \epsilon_s$, because in that case this k_z in the QW region will not be equal to k_z in the regions outside the well.)

Finally, using Eq. 5.21, repeated here for convenience:

$$\chi_{zz}(\omega, q_x; z, z') = \sum_\alpha \chi_\alpha(\omega, q_x)\xi_\alpha(z)\xi_\alpha(z') \tag{A.22}$$

we obtain the result of Ref. [101]:

$$\left(\partial_z^2 + k_z^2\right)D_z = -\sum_\alpha \frac{\chi_\alpha(\omega, q_x)}{\epsilon_s}\xi_\alpha(z)\int \xi_\alpha(z')\left(\partial_{z'}^2 + \epsilon_{xx}k_0^2\right)D_z(z')\,\mathrm{d}z' \tag{A.23}$$

A.2 A general solution

We can "guess" a solution to Eq. (A.23) of the following form:

$$D_z(z) = A\cos(k_z z) + B\sin(k_z z) + q_x^2 \sum_\alpha \frac{\chi_\alpha}{\epsilon_s}\left(AF_\alpha^A + BF_\alpha^B\right)\int \xi_\alpha(z')g(z,z')\,\mathrm{d}z' \tag{A.24}$$

with the Green's function $g(z,z') = -\sin(k_z|z-z'|)/2k_z$. This is quite similar to the solution in Ref. [101], however, their solution only works for adjacent transitions in symmetric quantum wells, where it can be assumed that $\xi_\alpha(z)$ is even with respect to z. This modified form will allow us to solve the equation without making such an assumption.

To see that this works as a solution, note that

$$\left(\partial_z^2 + k_z^2\right)\cos(k_z z) = 0, \tag{A.25}$$

$$\left(\partial_z^2 + k_z^2\right)\sin(k_z z) = 0, \tag{A.26}$$

$$\text{and} \quad \left(\partial_z^2 + k_z^2\right)g(z,z') = -\delta(z-z'). \tag{A.27}$$

So applying $\left(\partial_z^2 + k_z^2\right)$ to both sides of Eq. (A.24) gives us

$$\left(\partial_z^2 + k_z^2\right)D_z(z) = -q_x^2 \sum_\alpha \frac{\chi_\alpha}{\epsilon_s}\left(AF_\alpha^A + BF_\alpha^B\right)\xi_\alpha(z) \tag{A.28}$$

Comparing to (A.23), we see that our guess for $D_z(z)$ will be a solution if

$$AF_\alpha^A + BF_\alpha^B = \frac{1}{q_x^2}\int \xi_\alpha(z')\left(\partial_{z'}^2 + \epsilon_{xx}k_0^2\right)D_z(z')\,\mathrm{d}z' \tag{A.29}$$

At this point, our solution is self-referential, but it turns out that it is possible to solve for the F_α^A and F_α^B coefficients without knowing D_z first. (We will still be left with two unknowns A and B in the solution, but this is not surprising because we are searching for a general solution in the QW region, with unspecified boundary conditions. A and B will eventually be determined by the boundary conditions for a particular problem.)

A.3 Solving for the F coefficients

Note that, since $k_z^2 = \epsilon_{xx}k_0^2 - (\epsilon_{xx}/\epsilon_s)q_x^2$, we have

$$\left(\partial_z^2 + \epsilon_{xx}k_0^2\right)D_z = \frac{\epsilon_{xx}}{\epsilon_s}q_x^2 D_z + \left(\partial_z^2 + k_z^2\right)D_z \tag{A.30}$$

Using Eq. (A.28), this becomes

$$\left(\partial_z^2 + \epsilon_{xx}k_0^2\right)D_z = \frac{\epsilon_{xx}}{\epsilon_s}q_x^2 D_z - q_x^2\sum_\alpha \frac{\chi_\alpha}{\epsilon_s}\left(AF_\alpha^A + BF_\alpha^B\right)\xi_\alpha(z) \tag{A.31}$$

Then, plugging this into Eq. A.29 gives

$$AF_\alpha^A + BF_\alpha^B = \frac{\epsilon_{xx}}{\epsilon_s}\int \xi_\alpha(z')D_z(z')\,\mathrm{d}z' - \sum_{\alpha'}\frac{\chi_{\alpha'}}{\epsilon_s}\left(AF_{\alpha'}^A + BF_{\alpha'}^B\right)I_{\alpha,\alpha'}\,\mathrm{d}z \tag{A.32}$$

where we have defined for convenience later

$$I_{\alpha,\alpha'} = \int \xi_\alpha(z)\xi_{\alpha'}(z)\,\mathrm{d}z \tag{A.33}$$

Furthermore, by applying $(\epsilon_{xx}/\epsilon_s)\int \mathrm{d}z'\,\xi_\alpha(z')$ to Eq. (A.24), we see that

$$\frac{\epsilon_{xx}}{\epsilon_s}\int \xi_\alpha(z')D_z(z')\,\mathrm{d}z' = AC^A_\alpha + BC^B_\alpha - q_x^2\sum_{\alpha'}\frac{\chi_{\alpha'}}{\epsilon_s}\left(AF^A_{\alpha'} + BF^B_{\alpha'}\right)D_{\alpha,\alpha'} \tag{A.34}$$

where we have defined for convenience later

$$C^A_\alpha = \frac{\epsilon_{xx}}{\epsilon_s}\int \cos(k_z z)\xi_\alpha(z)\,\mathrm{d}z\,, \tag{A.35}$$

$$C^B_\alpha = \frac{\epsilon_{xx}}{\epsilon_s}\int \sin(k_z z)\xi_\alpha(z)\,\mathrm{d}z\,, \tag{A.36}$$

$$\text{and}\quad D_{\alpha,\alpha'} = -\frac{\epsilon_{xx}}{\epsilon_s}\int \xi_\alpha(z)g(z,z')\xi_{\alpha'}(z')\,\mathrm{d}z\,\mathrm{d}z'\,. \tag{A.37}$$

Plugging Eq. (A.34) into Eq. (A.32), we obtain

$$AF^A_\alpha + BF^B_\alpha = AC^A_\alpha + BC^B_\alpha - q_x^2\sum_{\alpha'}\frac{\chi_{\alpha'}}{\epsilon_s}\left(AF^A_{\alpha'} + BF^B_{\alpha'}\right)D_{\alpha,\alpha'} - \sum_{\alpha'}\frac{\chi_{\alpha'}}{\epsilon_s}\left(AF^A_{\alpha'} + BF^B_{\alpha'}\right)I_{\alpha,\alpha'} \tag{A.38}$$

Finally, we have an equation for the F_α's which does not depend self-referentially on the field $D_z(z)$. Our intention is that the A and B coefficients will depend on the boundary conditions, so we would like a solution for the F_α's that works for any and all values of A and B. As such, the A and B parts of the equation must be satisfied independently of each other, and we obtain the following two conditions:

$$F^A_\alpha = C^A_\alpha - \sum_{\alpha'}\frac{\chi_{\alpha'}}{\epsilon_s}F^A_{\alpha'}\left[I_{\alpha,\alpha'} + q_x^2 D_{\alpha,\alpha'}\right] \tag{A.39}$$

$$F^B_\alpha = C^B_\alpha - \sum_{\alpha'}\frac{\chi_{\alpha'}}{\epsilon_s}F^B_{\alpha'}\left[I_{\alpha,\alpha'} + q_x^2 D_{\alpha,\alpha'}\right] \tag{A.40}$$

Note that these can be written quite concisely as matrix equations:

$$\mathbf{M}\mathbf{F}^A = \mathbf{C}^A \tag{A.41}$$

$$\mathbf{M}\mathbf{F}^B = \mathbf{C}^B \tag{A.42}$$

with both equations using the same $\mathbf{M}$ matrix:

$$M_{\alpha,\alpha'} = \delta_{\alpha,\alpha'} + \frac{\chi_\alpha}{\epsilon_s}\left[I_{\alpha,\alpha'} + q_x^2 D_{\alpha,\alpha'}\right] \tag{A.43}$$

So, in summary, we have derived matrix equations for F^A_α, F^B_α which do not depend on $D_z(z)$. We can calculate $\mathbf{M}$, $\mathbf{C}^A$, and $\mathbf{C}^B$, only requiring knowledge of χ_α and $\xi_\alpha(z)$, which can be calculated from the Schrödinger-Poisson solution for the quantum well. Then, by solving the matrix equations, we obtain the F^A_α, F^B_α coefficients, which can be used to construct the $D_z(z)$ field solution (A.24). The two remaining unknowns A, B will be determined by the boundary conditions. So we could, for example, construct a transfer matrix similar to Ref. [101].

Finally, note that if $\xi_\alpha(z)$ is even (as would be the case for adjacent $n \to n+1$ transitions in a symmetric QW), then we will have $\mathbf{C}^B = 0$. In that case, we will always have $\mathbf{F}^B = 0$ as a valid solution. Then, if we also assume $\epsilon_{xx} = \epsilon_s$, we see that we recover the same solution as Ref. [101] in the case of inversion-symmetric quantum wells.

A.4 Effective permittivity

Now that we have a solution for D_z, we can try to construct an effective *local* permittivity tensor. We will do this by calculating the E_z field from D_z using Maxwell's equations, and then compare the two. As we will see, a local permittivity is only possible in the long-wavelength limit.

To find the E_z field from the D_z field, we can use (A.12), (A.13), along with $D_x = \epsilon_0 \epsilon_{xx} E_x$. After simplifying, we get

$$E_z(z) = \frac{1}{q_x^2 \epsilon_0 \epsilon_{xx}} \left(\partial_z^2 + \epsilon_{xx} k_0^2\right) D_z(z) \tag{A.44}$$

which, using $k_z^2 = \epsilon_{xx} k_0^2 - (\epsilon_{xx}/\epsilon_s) q_x^2$, can also be written as

$$E_z(z) = \frac{D_z(z)}{\epsilon_0 \epsilon_s} + \frac{1}{q_x^2 \epsilon_0 \epsilon_{xx}} \left(\partial_z^2 + k_z^2\right) D_z(z) \tag{A.45}$$

Using Eq. (A.28), we obtain an expression for the E_z field:

$$E_z(z) = \frac{D_z(z)}{\epsilon_0 \epsilon_s} - \frac{1}{\epsilon_0 \epsilon_{xx}} \sum_\alpha \frac{\chi_\alpha}{\epsilon_s} \left(A F_\alpha^A + B F_\alpha^B\right) \xi_\alpha(z) \tag{A.46}$$

Comparing this to our D_z field expression Eq. (A.24), it is clear (unsurprisingly) that we do not have a simple proportional relationship between the D_z and E_z fields with which to define a permittivity. However, we *can* define an effective permittivity in the long-wavelength limit by averaging the fields over z and taking a ratio $\langle D_z(z) \rangle \,/\, \langle E_z(z) \rangle$ (a similar approach to [61]). For terahertz (THz) QWs, the long-wavelength limit is quite reasonable, since we have well widths $\lesssim 100\,\text{nm}$ and electromagnetic wavelengths $\gtrsim 10\,\mu\text{m}$, even after accounting for the refractive index of the semiconductor.

In the long-wavelength limit, we take $q_x, k_0 \to 0$ (which also implies $k_z \to 0$). Then our expression (A.24) for the D_z field becomes simply

$$D_z(z) \approx A \tag{A.47}$$

We also have $C_\alpha^B \approx 0$ for all α, which means that $F_\alpha^B \approx 0$ as well. So our expression (A.46) for the E_z field becomes

$$E_z(z) \approx \frac{A}{\epsilon_0 \epsilon_s} - \frac{A}{\epsilon_0 \epsilon_{xx}} \sum_\alpha \frac{\chi_\alpha}{\epsilon_s} F_\alpha^A \xi_\alpha(z) \tag{A.48}$$

So we can define our effective permittivity (in the z direction) as

$$\boxed{\epsilon_{zz,\text{eff}} = \frac{\langle D_z \rangle}{\epsilon_0 \langle E_z \rangle} \approx \epsilon_s \left[1 - \sum_\alpha \frac{\chi_\alpha}{\epsilon_{xx}} F_\alpha^A \langle \xi_\alpha \rangle\right]^{-1}} \tag{A.49}$$

where $\langle \cdot \rangle$ denotes an average over the entire QW region in the z direction. E.g.,

$$\langle \xi_\alpha \rangle = \frac{1}{L} \int \xi_\alpha(z)\, \mathrm{d}z \tag{A.50}$$

where L is the thickness of the quantum well region. Furthermore, since we are using the long-wavelength limit anyway, we can also use it to simplify our equation for the F_α^A's:

$$F_\alpha^A \approx L \langle \xi_\alpha \rangle - \sum_{\alpha'} \frac{\chi_{\alpha'}}{\epsilon_s} F_{\alpha'}^A I_{\alpha,\alpha'} \tag{A.51}$$

So in the long wavelength limit, the problem is greatly simplified. Once we solve the matrix equation to find the F_α^A's, we can directly calculate an effective permittivity tensor which can be used in a standard electromagnetic solver.

As a final note, one must be slightly careful with the definition of L. Typically, L should be the width of the region simulated by Schrödinger-Poisson. So, for example, a common case might be a periodic stack of QWs separated by barriers. If one performs a Schrödinger-Poisson simulation of the entire stack, then L should be the length of the entire stack. If one performs a simulation of a single well with periodic boundary conditions, then L should be the length of the period. There is some arbitrariness here, but it is important to have consistency between the Schrödinger-Poisson equations, the calculation of the effective permittivity, and in the length of material which is assigned this effective permittivity in the final electromagnetic simulation.

Appendix B

Supplemental material on molecular beam epitaxy

This Appendix contains some additional comments and derivations related to the molecular beam epitaxy (MBE) chapters, Ch. 6 and Ch. 7. While not necessary to understand the main points of the chapters, these sections may be useful for certain readers.

B.1 Flux calibration coefficients

In Ch. 6, we defined three effusion cell calibration coefficients: α, β, and γ. However, in practice, the QNC-MBE lab uses a slightly different set of coefficients, A, B, and C, defined as

$$A = \frac{\alpha}{\gamma}; \quad B = \beta; \quad C = \gamma \tag{B.1}$$

This small tweak reflects the fact that the beam flux-monitoring ion gauge (BFM) versus temperature relation (Eq. (6.8)) is used more often than the flux versus temperature relationship (Eq. (6.4)) in day-to-day MBE growth, even though the latter is in some sense more fundamental.

As a result of this redefinition, the key MBE growth equations in practice become:

$$G = \frac{\left(a^{\parallel}\right)^2 a^{\perp}}{4} \cdot \phi \tag{B.2}$$

$$\phi = C I \sqrt{T} \tag{B.3}$$

$$I = \frac{A}{T} \exp\left(-\frac{B}{T}\right) \tag{B.4}$$

where, G is the growth rate and ϕ is the group III flux. We are assuming here that these are Zincblende III-V materials (4 group III atoms per unit cell) with the growth rate controlled exclusively by the group III flux. Also, note that it is important to use consistent units here. Quite often, flux is measured in $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ while growth rate and lattice constants are measured in terms of nanometres or angstroms. In that case, appropriate scaling factors must be applied.

B.2 Some useful approximations for the flux equations

In MBE, we run into several equations of the form

$$y = aT^{-r}\exp\left(-\frac{b}{T}\right) \tag{B.5}$$

In particular, we have the BFM-temperature relationship (Eq. (6.8)) with $r = 1$ and the flux-temperature relationship (Eq. (6.4)) with $r = 1/2$. As discussed in Sec. 6.4, we can invert such equations directly using the Lambert-W function. However, there are some approximations which are either useful in practice or useful for understanding sources of error.

B.2.1 Small temperature changes

Probably the most useful approximation is a simple Taylor series expansion about a temperature T_0:

$$y = aT^{-r}\exp\left(-\frac{b}{T}\right) \tag{B.6}$$

$$= aT_0^{-r}\exp\left(-\frac{b}{T_0}\right)\cdot\left[1 + \frac{b - rT_0}{T_0}\cdot\frac{T - T_0}{T_0} + O\left(\left(\frac{T - T_0}{T_0}\right)^2\right)\right] \tag{B.7}$$

If we define $y_0 = y(T = T_0)$, we can write this in an even simpler form:

$$\frac{y - y_0}{y_0} \approx \frac{b - rT_0}{T_0}\cdot\frac{T - T_0}{T_0} \tag{B.8}$$

So small changes in T are proportional to small *relative* changes in y. Notably, this proportionality does not depend on a, so we see that the b coefficient is largely the driver in terms of change versus temperature (at least over small temperature ranges).

To put this in more concrete terms, let us consider the specific case of flux. So $y = \phi$, $r = 1/2$, $a = \alpha$ and $b = \beta$. Then

$$\frac{\phi - \phi_0}{\phi_0} \approx \left(\frac{\beta}{T_0} - \frac{1}{2}\right)\frac{T - T_0}{T_0} \tag{B.9}$$

This relationship is useful in practice during flux calibrations. If we are close to the desired flux, but not quite there, we can invert this relationship to determine how much the cell temperature needs to be adjusted to reach the desired flux. Typically, the approximation will be plenty accurate as long as the temperature change is not more than a few degrees or so.

We can gain some intuition for these changes by simplifying even further. β is typically on the order of 10 000 K while T is around 1000 K, so we can neglect the $1/2$ to write

$$\frac{\phi - \phi_0}{\phi_0} \approx \frac{\beta}{T_0^2}(T - T_0) \tag{B.10}$$

This gives us a very nice rule of thumb for cell temperature changes. If we assume $T_0 \sim 1000\,\mathrm{K}$, then the flux will change by $(\beta/10000)\%$ for every 1 K change in cell temperature. For example, if a cell has $\beta \sim 30\,000\,\mathrm{K}$, then we know that every 1 degree change in cell temperature will lead to around a 3 % change in flux. Thus, we can very quickly gain an intuition for how a cell responds to changes in temperature by looking at its β coefficient.

B.2.2 Eliminating square roots of temperature

Although expressions of the form Eq. B.5 can be inverted with the Lambert W function, MBE growers sometimes make things easier by dropping the factor of T^{-r} entirely, writing

$$y \approx a \exp\left(-\frac{b}{T}\right) \tag{B.11}$$

At first glance, this may appear difficult to justify – the units/dimensions do not even match the original equation. And yet, the approximation tends to work quite well in practice, so it is interesting to explore why. In fact, the apparent lack of dimensional consistency is a clue as to what's going on. To fix the dimensional consistency, we might guess that there has been some implicit redefinition of the a coefficient. These coefficients are simply fitted numerical parameters in practice, after all. If we *also* allow for the possibility that the b coefficient has been redefined, then it turns out that there is some validity to this approximation. However, it is still an approximation, not an exact relationship, and it will only work for a limited range of temperatures. As we deviate further from the temperatures used for calibration, the approximation will start to fail.

To understand this, assume T is close to some typical calibration temperature T_0. Specifically, let us assume that we can write

$$\frac{T_0}{T} = 1 + \epsilon \tag{B.12}$$

with $|\epsilon| \ll 1$. Since ϵ is small, we can make use of the two following Taylor series expansions:

$$(1+\epsilon)^r = 1 + r\epsilon + O\left(\epsilon^2\right) \tag{B.13}$$

$$\exp(r\epsilon) = 1 + r\epsilon + O\left(\epsilon^2\right) \tag{B.14}$$

We see that these two expressions are equivalent up to second-order in $r\epsilon$, so we can write

$$(1+\epsilon)^r = \exp(r\epsilon) + O\left(\epsilon^2\right) \tag{B.15}$$

Converting back to temperatures, we can therefore write

$$\left(\frac{T_0}{T}\right)^r = \exp\left(r\left(\frac{T_0}{T} - 1\right)\right) + O\left(\left(r\left(\frac{T_0 - T}{T}\right)\right)^2\right) \tag{B.16}$$

Rearranging, and dropping terms of second order, we obtain

$$T^{-r} \approx T_0^{-r} e^{-r} \exp\left(r\frac{T_0}{T}\right). \tag{B.17}$$

Finally, we can plug this approximation into Eq. (B.5) to obtain

$$y \approx a T_0^{-r} e^{-r} \exp\left(r\frac{T_0}{T}\right) \exp\left(-\frac{b}{T}\right) \tag{B.18}$$

$$= a T_0^{-r} e^{-r} \exp\left(-\frac{b - rT_0}{T}\right) \tag{B.19}$$

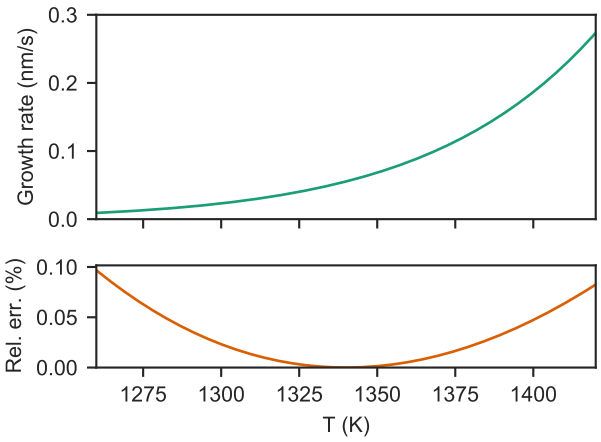

Figure B.1: Example of the second-order error generated by eliminating the $\sqrt{T}$ term from the flux equation. The top panel shows the growth rate of AlAs on GaAs versus Al cell temperature, calculated using typical values of calibration coefficients (described in the main text). The bottom panel shows the relative error when performing the same calculation without the $\sqrt{T}$ term in the flux equation.

If we define a new set of coefficients

$$a' = aT_0^{-r}e^{-r}; \quad b' = b - rT_0 \tag{B.20}$$

then we obtain the simple form

$$y \approx a' \exp\left(-\frac{b'}{T}\right) \tag{B.21}$$

which will be accurate to second order in $(T_0 - T)/T$.

A simple example is plotted in Fig. B.1, using typical values from our Al cell. The Al flux is calculated from the cell temperature using $\phi = aT^{-1/2}\exp(-b/T)$ with $a = 1.56 \times 10^{14}\,\mathrm{K^{1/2}\,nm^{-2}\,s^{-1}}$ and $b = 38\,700\,\mathrm{K}$. This is converted into an AlAs growth rate, assuming that it is grown epitaxially on a GaAs substrate. An alternative growth rate is calculated using $\phi = a'\exp(-b'/T)$ with a', b' calculated from a, b according to Eq. (B.20), using $T_0 = 1340\,\mathrm{K}$ (the temperature in the centre of the plot range). The relative error between these two modes of calculation is plotted in the bottom panel, and we see that it is smaller than 0.1 % over a typical range of growth rates 0.01–0.30 nm/s. The error has a clear quadratic dependence, reaching zero at T_0, as we have predicted.

So, similar to what we saw for γ in Sec. 6.6, if an MBE grower neglects the factors $1/\sqrt{T}$ and $1/T$ in their flux-temperature (Eq. (6.4)) and BFM-temperature (Eq. (6.8)) relationships, the equations will work to a reasonable approximation, thanks to an implicit redefinition of the calibration coefficients. While this neglect of the factors is sometimes done in practice, it should be stressed that it is not actually necessary, since the MBE growth equations can be inverted to arbitrary precision via the Lambert-W function, as outlined in Sec. 6.4.

B.2.3 Approximation for temperature offset

One thing that *cannot* be eliminated in practice are errors in the temperature reading. As discussed in Sec. 6.6, we are fortunate that such errors in the BFM-flux relationship (Eq. (6.7)) are approximately cancelled out due to an implicit redefinition of the calibration coefficient γ. A similar thing happens for the BFM-temperature relationship (Eq. (6.8)) and the flux-temperature relationship (Eq. (6.4)), which we will demonstrate here.

We will do this by demonstrating that the following approximation holds to second order in variations in T:

$$\frac{a}{(T+\Delta T)^r}\cdot\exp\left(-\frac{b}{T+\Delta T}\right)\approx\frac{a'}{T^r}\exp\left(-\frac{b'}{T}\right) \tag{B.22}$$

In other words, shifting our measurement T by ΔT is approximately equivalent to a redefinition of the coefficients $a\to a'$ and $b\to b'$. To see how this works, let us expand both sides of the equation in Taylor series around a reference temperature T_0. First, the left-hand side:

$$\begin{aligned}&\frac{a}{(T+\Delta T)^r}\cdot\exp\left(-\frac{b}{T+\Delta T}\right)=\\&\quad\frac{a}{(T_0+\Delta T)^r}\exp\left(-\frac{b}{T_0+\Delta T}\right)\cdot\left[1+\frac{b-r(T_0+\Delta T)}{(T_0+\Delta T)^2}\cdot(T-T_0)\right]+O\left(\frac{(T-T_0)^2}{T_0^2}\right)\end{aligned} \tag{B.23}$$

And then the right-hand side:

$$\frac{a'}{T^r}\cdot\exp\left(-\frac{b'}{T}\right)=\frac{a'}{T_0^r}\exp\left(-\frac{b'}{T_0}\right)\cdot\left[1+\frac{b'-rT_0}{T_0^2}\cdot(T-T_0)\right]+O\left(\frac{(T-T_0)^2}{T_0^2}\right) \tag{B.24}$$

Then the idea is to choose a' and b' so that these two expressions to be equal up to second order. In other words, we want the zeroth and first-order terms of the Taylor expansions to be identical. This procedure gives use the following two conditions, from which we can extract a' and b':

$$\frac{a}{(T_0+\Delta T)^r}\exp\left(-\frac{b}{T_0+\Delta T}\right)=\frac{a'}{T_0^r}\exp\left(-\frac{b'}{T_0}\right) \tag{B.25}$$

$$\frac{b-r(T_0+\Delta T)}{(T_0+\Delta T)^2}=\frac{b'-rT_0}{T_0^2} \tag{B.26}$$

We can make these expressions a bit more manageable by defining $\xi=T_0/(T_0+\Delta T)$. Then the conditions become

$$\xi^r a\exp\left(-\frac{\xi b}{T_0}\right)=a'\exp\left(-\frac{b'}{T_0}\right) \tag{B.27}$$

$$\xi^2 b-\xi rT_0=b'-rT_0 \tag{B.28}$$

Noting that $1-\xi=\xi\Delta T/T_0$, we can show that the solution for b' is:

$$b'=\xi^2 b+\xi r\Delta T \tag{B.29}$$

and then the solution for a' is

$$a' = \xi^r a \exp\left[\frac{\xi \Delta T}{T_0} \cdot \left(-\frac{\xi b}{T_0} + r\right)\right]. \tag{B.30}$$

So we see that it is indeed possible to solve for a' and b' such that the zeroth and first-order error terms drop out. With a', b' defined like this, we can write our approximation in a more precise way:

$$\frac{a}{(T+\Delta T)^r} \cdot \exp\left(-\frac{b}{T+\Delta T}\right) \approx \frac{a'}{T^r} \exp\left(-\frac{b'}{T}\right) + O\left(\left(\frac{T-T_0}{T_0}\right)^2\right) \tag{B.31}$$

In other words, we have shown that a constant temperature reading offset is approximately equivalent to a redefinition of the calibration coefficients, as long as we work in a limited temperature range near some T_0. To give an example of numbers, we can use the same example as Sec. B.2.2, growing AlAs on GaAs with typical Al cell parameters. If we assume a temperature offset of $\Delta T = 50\,\mathrm{K}$, then the error varies quadratically in the range 0.0–0.5 % for growth rates in the range 0.01–0.3 nm/s So, similar to what we saw in Sec. 6.6, if we perform a calibration procedure in the presence of a constant temperature error ΔT, the calibration will tend to choose the coefficients a' and b' that mitigate this error.

B.3 Finding the transfer function with Bode plots

In Sec. 7.4, we presented the following second-order approximation of the Al cell transfer function:

$$H(f) = \frac{T_{\mathrm{eff}}(f)}{T_{\mathrm{in}}(f)} \approx \frac{K\zeta \cdot (i2\pi f\tau) + 1}{(i2\pi f\tau)^2 + 2\zeta \cdot (i2\pi f\tau) + 1} \tag{B.32}$$

This particular form of the transfer function works for our Al cell, but it might not work in general. It may be useful, therefore, to understand where this expression came from in the first place. While there was no rigorous process followed here, there are some useful techniques which can help guide our choice of transfer function.

The first step is to perform a direct calculation of the transfer function by straight division in the frequency domain. After applying some known input $T_{\mathrm{in}}(t)$ and measuring a response $T_{\mathrm{eff}}(t)$, one takes Fourier transforms and performs the division $H(f) = T_{\mathrm{eff}}(f)/T_{\mathrm{in}}(f)$. The results of this are shown in Figs. B.2 and B.3. Already, one needs to be careful here, as real-world data will be discrete, which brings additional challenges in the Fourier transform. As good practice, one should make sure that the data is evenly-spaced in the time domain, multiplied by an appropriate window function, and zero-padded as necessary. More important still is the choice of input function, $T_{\mathrm{in}}(t)$, which should be chosen so that it does not have any zeros (or near-zeros) in the frequency domain – otherwise one ends up with numerical problems from dividing by small numbers.

With these caveats, in principle, we could already use this simple frequency domain division to calculate our transfer function. In fact, this is exactly the technique that was used in Ref. [115]. However, calculating the transfer function in this way is somewhat crude, incorporating noise from the measurement and being difficult to work with. Furthermore, there is a danger in fitting the results too closely. For example, the response shown in B.2 is not symmetric – the response to a step-up is slightly different from the response to a step-down. We expect that this will be a general behaviour that that does not follow the rules of linear systems. So, while we may be able to find a linear transfer function that reproduces the effect for this

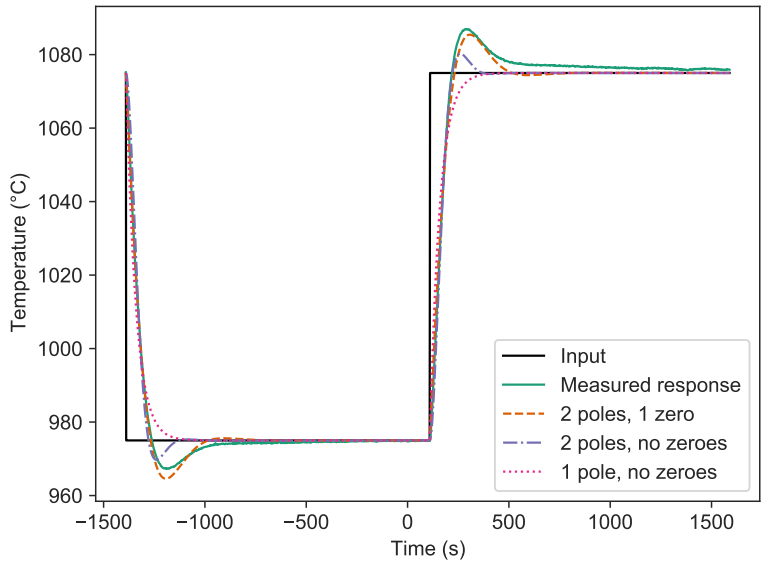

Figure B.2: Similar to Fig. 7.1, this shows the step input $T_{\text{in}}(t)$ (black) and the $T_{\text{eff}}(t)$ response (green). The dotted lines show the least-squares fits from several possible models with different poles and zeros. We can see that at least two poles and one zero are necessary to capture the key features of the response.

particular input, we cannot expect to find a transfer function that will reproduce it for all inputs. And, in fact, by trying to force a perfect fit for this particular input, we may make things worse for a general input.

We would prefer, instead, to come up with a simple form of the transfer function with a few fit-able parameters. To figure out what form this simplified transfer function might take, we can use a classic electrical engineering tool: the Bode plot. A Bode plot of the transfer function is given in Fig. B.3. Bode plots allow use to quickly understand a transfer function in terms of its poles and zeros. We could write a general transfer function of the form:

$$H(f) = C\frac{(1+if/z_0)\cdot(1+if/z_1)\cdot(1+if/z_2)\cdots}{(1+if/p_0)\cdot(1+if/p_1)\cdot(1+if/p_2)\cdots} \tag{B.33}$$

where z_i are called the zeros of the transfer function and p_i are called the poles. Roughly, each zero and pole will cause a "kink" in the Bode plot of the transfer function at a certain frequency. (If the pole or zero is real, this kink happens exactly at $f = z_i$ or $f = p_i$.) Above this critical frequency, a zero will cause the magnitude to slope upward by 20 dB/decade, and it will cause the phase to rise by 90°. Conversely, a pole will cause the magnitude to drop by −20 dB/decade, and it will cause the phase to drop by 90°.

Looking the Bode plot of our transfer function in Fig. B.3, we can see that the magnitude is flat with a value of 0 dB up to around 2×10^{-3} Hz, after which it drops with a slope of around −20 dB/decade. At a similar frequency, we see the phase drop from around 0° to −90°. This is suggestive of a pole around 2×10^{-3} Hz. However, in the time-domain version of the step response in Sec. 7.4, we see an overshoot behaviour, which is indicative of at least a second-order system. So, another possibility is that there are

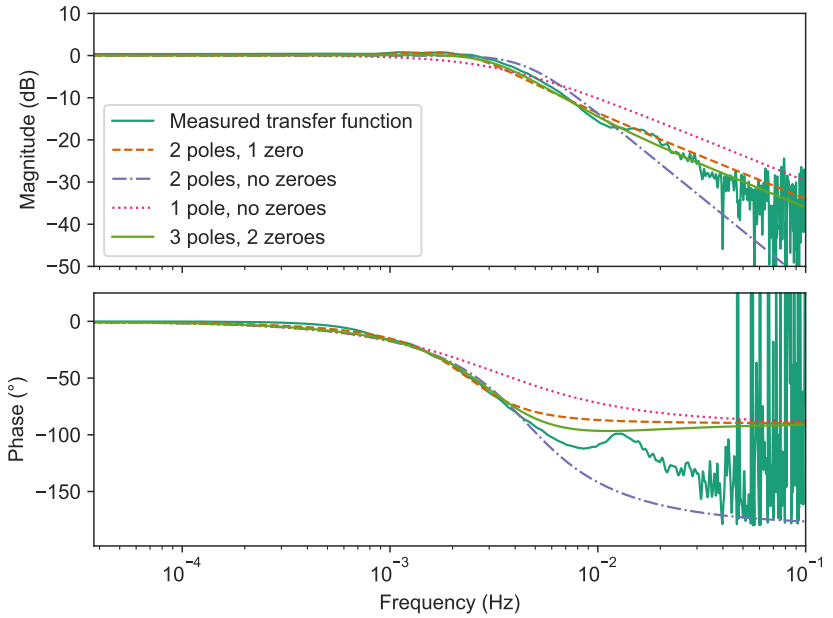

Figure B.3: A Bode plot of the measured transfer function (green), which comes from the raw calculation $T_{\text{eff}}(f)/T_{\text{in}}(f)$. The dotted lines show the best fits from several simple transfer function models. The top panel shows the magnitude in dB (i.e., $20\log_{10}|H(f)|$) and the bottom panel shows the phase, $\angle H(f)$.

two poles and one zero all giving "kinks" at a similar frequency, so that the phase and slope behaviour of one pole are counteracted by the zero. As such, we can guess that the transfer function takes a general two-pole one-zero form with parameters to fit. This is exactly what the transfer function of Eq. (B.32) represents, although we have not written it in the explicit pole-zero form of Eq. (B.33). Further, we have restricted the coefficients somewhat so that $H(0) = 1$, to ensure that the steady-state values of $T_{\text{eff}}(t)$ and $T_{\text{in}}(t)$ will be equal.

As a comparison, Figs. B.2 and B.3 show what happens if we use fewer poles or zeros. In both the frequency domain and the time domain, we can see that the approximation improves as more zeros and poles are added. Of course, we could improve the fit by adding even more poles and zeros, but, again, one must be careful of overfitting a linear model to a system that is known to be non-linear. Overall, then, the idea is to add just enough poles and zeros to capture the key dynamics.

Appendix C

Mathematical model of multipass absorption

In this appendix we derive an expression for multipass absorption through a quantum well. This approach was used to fit the experimental data in Sec. 8.1. As illustrated in Fig. 8.2, we consider a sample with three key regions: a substrate, an active region, and a metallic coating. For this model, we treat the metal as a perfect electric conductor, and the substrate as having a constant, real permittivity ϵ_{sub}. We model the active region with an effective complex permittivity tensor. For this work, where the active region comprises a stack of quantum wells, we assume that the growth-direction component ϵ_{zz} takes the form of Eq. 8.1, and that $\epsilon_{xx} = \epsilon_{yy} = \epsilon_s$. However, the electromagnetic derivation that follows is general and does not depend on the particular form of the permittivity tensor, so more sophisticated models could be used.

The main restriction here is that each individual quantum well must be small compared to the wavelength of light, so that the stack can be treated as an effective medium from the perspective of the electromagnetic field. In the case of Ch. 8.1, we are working with free-space wavelengths on the order of 100 μm, which will give a growth-direction wavelength of around $\lambda \approx 40\,\mu\text{m}$ in GaAs at 45° incidence. So, the individual quantum wells, which have a width on the order of 0.1 μm, are quite small compared to the wavelength.

It should be emphasized, though, that this does not mean the stack as a whole is negligibly thin. In the case of the 54-well stack, the total thickness of the active region is around 6 μm, which is an appreciable fraction of the wavelength. On the other extreme, the substrate thickness (∼500 μm) is quite large compared to the wavelength, and we will neglect it: effectively assuming that the distance is too far to achieve coherent reflections. Because of this, as the light bounces back and forth between the two surfaces of the sample (again, see Fig. 8.2), we can treat each "pass" or reflection through the active region independently. Thus, if we can find the power absorption spectrum for a single pass (which we will do in Sec. C.1), it will be a simple matter to extend this based on the total number of passes to find the total absorption through the structure (which we will do in Sec. C.2).

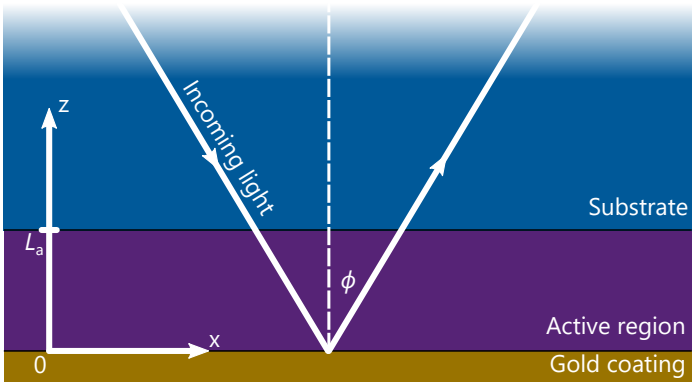

Figure C.1: Geometry used to derive the single-pass reflection coefficient. Note: to ensure a right-handed coordinate system, the y axis points into the page.

C.1 The single-pass reflection coefficient

We will now calculate the complex reflection coefficient for a single reflection through the active region. See Fig. C.1 for an illustration of the coordinates we will use. We let z be the growth direction, and let the active region lie in $0 < z < L_a$ with a perfect conductor for $z < 0$. At $z > L_a$ there is substrate with permittivity ϵ_{sub}. (We assume that the light arrives at the active region from far away in the substrate, so we can treat the substrate as a semi-infinite slab.)

For TM light, let's assume the magnetic field, $\mathbf{H}$, is pointing in the y direction. Then $\mathbf{H}$ can be written in the active region and substrate, respectively, as

$$H_{y,a} = \left(H_a^- e^{-ik_{z,a}z} + H_a^+ e^{+ik_{z,a}z}\right) e^{ik_{x,a}x}; \quad 0 < z < L_a \tag{C.1}$$

$$\text{and } H_{y,s} = \left(H_s^- e^{-ik_{z,s}(z-L_a)} + H_s^+ e^{+ik_{z,s}(z-L_a)}\right) e^{ik_{x,s}x}; \quad z > L_a. \tag{C.2}$$

Using Maxwell's equations, we have

$$i\omega\epsilon_0(\epsilon \cdot \mathbf{E}) = \boldsymbol{\nabla} \times \mathbf{H} = \left(\partial_x H_y \hat{\mathbf{z}} - \partial_z H_y \hat{\mathbf{x}}\right) \tag{C.3}$$

So, in the active region, we have

$$E_{z,a} = \frac{k_{x,a}}{\omega\epsilon_{zz}\epsilon_0}\left(H_a^- e^{-ik_{z,a}z} + H_a^+ e^{+ik_{z,a}z}\right) e^{ik_{x,a}x} \tag{C.4}$$

$$E_{x,a} = \frac{k_{z,a}}{\omega\epsilon_{xx}\epsilon_0}\left(H_a^- e^{-ik_{z,a}z} - H_a^+ e^{+ik_{z,a}z}\right) e^{ik_{x,a}x} \tag{C.5}$$

and in the substrate region, we have

$$E_{z,s} = \frac{k_{x,s}}{\omega \epsilon_{\text{sub}} \epsilon_0} \left(H_s^- e^{-ik_{z,s}(z-L_a)} + H_s^+ e^{+ik_{z,s}(z-L_a)} \right) e^{ik_{x,s}x} \tag{C.6}$$

$$E_{x,s} = \frac{k_{z,s}}{\omega \epsilon_{\text{sub}} \epsilon_0} \left(H_s^- e^{-ik_{z,s}(z-L_a)} - H_s^+ e^{+ik_{z,s}(z-L_a)} \right) e^{ik_{x,s}x} \tag{C.7}$$

Because of the conductor boundary condition, we must have $E_{x,a} = 0$ at $z = 0$, which gives

$$\frac{k_{z,a}}{\omega \epsilon_{xx} \epsilon_0} \left(H_a^- - H_a^+ \right) e^{ik_{x,a}x} = 0 \tag{C.8}$$

or $H_a^- = H_a^+ := H_a$. So the fields in the active region are given by

$$H_{y,a} = 2H_a \cos\left(k_{z,a} z\right) e^{ik_{x,a}x} \tag{C.9}$$

$$E_{z,a} = 2H_a \frac{k_{x,a}}{\omega \epsilon_{zz} \epsilon_0} \cos\left(k_{z,a} z\right) e^{ik_{x,a}x} \tag{C.10}$$

$$E_{x,a} = -2iH_a \frac{k_{z,a}}{\omega \epsilon_{xx} \epsilon_0} \sin\left(k_{z,a} z\right) e^{ik_{x,a}x} \tag{C.11}$$

Then, our second boundary condition is that the tangential **E** and **H** fields must be continuous across the boundary at $z = L_a$. First, this enforces that $k_{x,s} = k_{x,a}$. It also gives

$$2H_a \cos\left(k_{z,a} L_a\right) = \left(H_s^- + H_s^+ \right) \tag{C.12}$$

$$-2iH_a \frac{k_{z,a}}{\omega \epsilon_{xx} \epsilon_0} \sin\left(k_{z,a} L_a\right) = \frac{k_{z,s}}{\omega \epsilon_{\text{sub}} \epsilon_0} \left(H_s^- - H_s^+ \right) \tag{C.13}$$

Eliminating H_a, we get

$$\frac{\omega \epsilon_{xx} \epsilon_0}{-ik_{z,a}} \cot\left(k_{z,a} L_a\right) = \frac{\omega \epsilon_{\text{sub}} \epsilon_0}{k_{z,s}} \frac{H_s^- + H_s^+}{H_s^- - H_s^+} \tag{C.14}$$

which becomes

$$i \frac{\epsilon_{xx} k_{z,s}}{k_{z,a} \epsilon_{\text{sub}}} \cot\left(k_{z,a} L_a\right) = \frac{1 + r}{1 - r} \tag{C.15}$$

Where $r = H_s^+ / H_s^-$ is the complex reflection coefficient. Solving for r, we find that the complex reflection coefficient is

$$\boxed{r = \frac{-k_{z,a} \epsilon_{\text{sub}} + ik_{z,s} \epsilon_{xx} \cot\left(k_{z,a} L_a\right)}{k_{z,a} \epsilon_{\text{sub}} + ik_{z,s} \epsilon_{xx} \cot\left(k_{z,a} L_a\right)}} \tag{C.16}$$

To actually calculate r, we still need to know the wavenumbers in the substrate and active region. The wavenumbers in the substrate region must obey

$$k_{x,s}^2 + k_{z,s}^2 = \epsilon_{\text{sub}} \frac{\omega^2}{c^2} = \epsilon_{\text{sub}} k_0^2, \tag{C.17}$$

so we can write [1]

$$k_{x,s} = \sqrt{\epsilon_{\text{sub}}} k_0 \sin\phi; \quad k_{z,s} = \sqrt{\epsilon_{\text{sub}}} k_0 \cos\phi \tag{C.18}$$

where, ϕ is the incidence angle, measured relative to the growth axis.

The wavenumbers in the active region are a bit trickier because of the anisotropy, but it can be shown (using $\boldsymbol{\nabla} \times \mathbf{E} = -i\omega\mu_0 \mathbf{H}$) that

$$\frac{k_{x,a}^2}{\epsilon_{zz}} + \frac{k_{z,a}^2}{\epsilon_{xx}} = \frac{\omega^2}{c^2} = k_0^2 \tag{C.19}$$

Recall from before that we have $k_{x,a} = k_{x,s}$. So, solving for $k_{z,a}$ in terms of k_0, we get

$$k_{z,a}^2 = \epsilon_{xx}\left[1 - \frac{\epsilon_{\text{sub}}}{\epsilon_{zz}} \sin^2\phi\right] k_0^2 \tag{C.20}$$

We now have expressions for $k_{z,s}$ and $k_{z,a}$, which we can use in Eq. C.16 to calculate the complex reflection coefficient for the active region against a gold coating.

C.2 Multipass absorption

In a multipass configuration (again, where we are assuming incoherent reflections off the bottom of the substrate so that each reflection is independent), the total transmission through the structure can be calculated as

$$T(\omega) = |r(\omega)|^{2N_{\text{pass}}} \tag{C.21}$$

where N_{pass} is the number of passes through the active region (i.e., the number of reflections). This can be approximately calculated [90] using

$$N_{\text{pass}} = \frac{L_{\text{sub}}}{D_{\text{sub}}} \cot\phi \tag{C.22}$$

where L_{sub} and D_{sub} are the length and thickness of the sample, respectively, and ϕ is the incidence angle. However, this is only an approximation. In reality, the number of passes may depend on where the light hits the facet. So, in general, one would need to know the distribution of light across the facet to calculate the total transmission. In practice, it is quite difficult to know this value exactly, which translates to an uncertainty in the total magnitude of the absorption.

Further, note that we have neglected the reflection off the entry and exit facets: i.e., only a certain fraction of the incoming light will actually enter the sample in the first place. In practice, however (as mentioned in Sec. 8.1) such complications are mitigated by dividing the TM absorption measurement by an identical TE measurement. This cancels out any losses which are common to both TM and TE light, isolating the effect of the active region, which is anisotropic and only absorbs TM light.

[1] Note we have implicitly assumed $k_{x,s}$ and $k_{z,s}$ to be real here, since we are assuming that the other substrate boundary is far off and thus the incoming light can be treated as a plane wave. If the substrate thickness were not very large compared to the wavelength, we could not necessarily assume that $k_{x,s}$ and $k_{z,s}$ are real, and we would need to consider waveguide modes.

C.3 The absorption coefficient

In Sec. 5.5, we discussed the notion of an absorption coefficient. The absorption coefficient is also commonly used by the community in multipass experiments (e.g., Ref. [90]), and it is interesting to confirm that it falls out of the theory we have just developed – at least, under certain approximations. In particular, we will need to assume that the active region is small compared to the wavelength, and that the absorption per pass is not too strong. We will also assume that $\epsilon_{xx} = \epsilon_{\text{sub}}$. In that case, the complex reflection coefficient of Eq. C.16 can be written as

$$r = \frac{1 + i\left(k_{z,a}/k_{z,s}\right)\tan\left(k_{z,a}L_a\right)}{1 - i\left(k_{z,a}/k_{z,s}\right)\tan\left(k_{z,a}L_a\right)} \tag{C.23}$$

Assuming $k_{z,a}L_a$ is small, we have $\tan\left(k_{z,a}L_a\right) \approx k_{z,a}L_a$, and so

$$r \approx \frac{1 + iL_a k_{z,a}^2/k_{z,s}}{1 - iL_a k_{z,a}^2/k_{z,s}} \tag{C.24}$$

If we assume that $L_a k_{z,a}^2/k_{z,s}$ is small, we can simplify this to

$$r \approx 1 + 2i\frac{k_{z,a}^2 L_a}{k_{z,s}} \tag{C.25}$$

We can then obtain the power spectrum of the reflection, keeping only terms to first order:

$$|r|^2 = r \cdot r^* \tag{C.26}$$

$$\approx 1 + 2i\left(k_{z,a}^2 - \left(k_{z,a}^2\right)^*\right) \cdot \frac{L_a}{k_{z,s}} \tag{C.27}$$

$$= 1 - 4\text{Im}\left[k_{z,a}^2\right] \cdot \frac{L_a}{k_{z,s}} \tag{C.28}$$

Using our expressions for $k_{z,a}^2$ and $k_{z,s}$ from Sec. C.1, this becomes

$$|r|^2 \approx 1 + 4\sqrt{\epsilon_{\text{sub}}}k_0 L_a \,\text{Im}\left[\frac{\epsilon_{\text{sub}}}{\epsilon_{zz}(\omega)}\right]\frac{\sin^2\phi}{\cos\phi} \tag{C.29}$$

If we define the dimensionless absorption coefficient as

$$\alpha_{2\text{D}}(\omega) = -\frac{\omega\sqrt{\epsilon_s}L_p}{c}\,\text{Im}\left[\frac{\epsilon_{\text{sub}}}{\epsilon_{zz}(\omega)}\right] \tag{C.30}$$

where L_p is the width of a single quantum well (QW) period in the active region, then we obtain

$$|r|^2 \approx 1 - 4\frac{\sin^2\phi}{\cos\phi}N_p\alpha_{2\text{D}}(\omega) \tag{C.31}$$

where $N_p = L_a/L_p$ is the total number of periods in the active region. This can also be approximated as

$$|r|^2 \approx \exp\left(-4\frac{\sin^2\phi}{\cos\phi}L_p\alpha_{2D}(\omega)\right) \tag{C.32}$$

which is the standard exponential-decay form used in Eq. (39) of Ref. [90].[2] Note, of course, that we had to assume a small active region compared to the wavelength ($k_{z,a}L_a \ll 1$) to obtain this formula. Less obvious is that we had to implicitly assume weak absorption in the QW active region. Weak absorption is required because we have assumed that $k_{z,a}L_a$ and $k_{z,a}^2 L_a/k_{z,s}$ are small, but $k_{z,a}$ depends on $1/\epsilon_{zz}$ (Eq. (C.20)). If the QW absorption is strong, then the imaginary part of $1/\epsilon_{zz}$ will become large, making $k_{z,a}$ large and breaking our approximation. Nevertheless, we have seen that in the limit of a thin active region and weak absorption, the power follows a simple exponential decay rule with an absorption coefficient $\alpha_{2D}(\omega)$.

C.4 Integrated absorption

It is often useful to quantify the total absorption strength of a QW active region. In the case of Ch. 8.1, we have done this using the plasma frequency ω_p as a measure of absorption strength. It is not necessarily obvious that this is a good quantifier of absorption strength though. One alternative way to measure the absorption strength is by looking at the integral of the absorption coefficient:

$$I = \int_0^\infty \alpha_{2D}(\omega)\,\mathrm{d}\omega \tag{C.33}$$

This has a clearer interpretation, but can only be used in cases where the absorption coefficient is well-defined. What we will show in this section is that these two quantifiers of absorption strength are essentially equivalent – at least in cases where both are well-defined. We will see that I turns out to be equal to ω_p^2, aside from some scaling factors.

We want to evaluate the integrated absorption analytically, so first we must find an expression for the absorption coefficient, (C.30), using the single-mode form of the dielectric tensor, Eq. (8.1). We have

$$\alpha_{2D}(\omega) = -\frac{\omega\sqrt{\epsilon_s}L_p}{c}\,\mathrm{Im}\left[1 + \frac{\omega_p^2}{\omega^2 - \omega_0^2 + i\gamma\omega}\right] \tag{C.34}$$

which simplifies to

$$\alpha_{2D}(\omega) = \frac{\sqrt{\epsilon_s}L_p}{c}\frac{\gamma\omega_p^2\omega^2}{\left(\omega^2-\omega_0^2\right)+(\gamma\omega)^2} \tag{C.35}$$

Thus, the integrated absorption is given by

$$I = \frac{\omega_p^2\sqrt{\epsilon_s}L_p}{c}\int_0^\infty \frac{\gamma\omega^2}{\left(\omega^2-\omega_0^2\right)+(\gamma\omega)^2}\,\mathrm{d}\omega \tag{C.36}$$

[2] In our case, the number of passes is equal to 1 ($M = 2$ in Ref. [90]), and the overlap factor is maximized at 2, since we have assumed a thin active region next to a perfect conductor.

We can simplify the integral through the change of variables $u = \omega_0/\omega$ ($\implies \mathrm{d}\omega = -\omega_0\,\mathrm{d}u\,/u^2$) and $G = \gamma/2\omega_0$. We obtain

$$I = \frac{\omega_p^2\sqrt{\epsilon_s}L_p}{c}\int_0^\infty f(u)\,\mathrm{d}u\,; \qquad f(u) = \frac{2G}{(u^2-1)^2+4G^2u^2} \tag{C.37}$$

Note that, since $f(u)$ is even, we can write

$$\int_0^\infty f(u)\,\mathrm{d}u = \frac{1}{2}\int_{-\infty}^\infty f(u)\,\mathrm{d}u \tag{C.38}$$

Now, if we treat $f(u)$ as a complex function, we can use the residue technique to evaluate its integral [136]. In the complex plane, our integral from $-\infty$ to ∞ corresponds to a path integral along the entire real line. We can turn this into a closed path, C, by adding a semicircle of infinite radius in the upper half of the complex plane (see Fig. C.2). This semicircular part of the path will contribute nothing to the integral because $f(|u|\to\infty) = 0$. In other words:

$$\int_{-\infty}^\infty f(u)\,\mathrm{d}u = \oint_C f(u)\,\mathrm{d}u \tag{C.39}$$

which means that our integrated absorption is given by

$$I = \frac{\omega_p^2\sqrt{\epsilon_s}L_p}{2c}\oint_C f(u)\,\mathrm{d}u \tag{C.40}$$

Expressing this in terms of a closed path integral is advantageous because we can now use the residue theorem to evaluate it. The residue theorem says

$$\oint_C f(u)\,\mathrm{d}u = 2\pi i\sum_p \mathrm{Res}(f,p) \tag{C.41}$$

where $\mathrm{Res}(f,p)$ is the residue of f at p, and p sums over all the poles of f enclosed by C: i.e., all the poles in the upper half of the complex plane. So we just need to find the poles and evaluate the residues. To find the poles, we can begin to factorize the denominator of $f(u)$:

$$f(u) = \frac{2G}{(u^2-1+i2Gu)(u^2-1-i2Gu)} \tag{C.42}$$

Then, if we define $H = \sqrt{1-G^2}$, we can complete the factorization to obtain

$$f(u) = \frac{2G}{(u+iG-H)(u+iG+H)(u-iG-H)(u-iG+H)} \tag{C.43}$$

So we have four poles at $iG\pm H$ and $-iG\pm H$. The poles at $iG\pm H$ will always be in the upper half of the complex plane (i.e., they have positive imaginary part), while the poles at $-iG\pm H$ will always be in the

lower half (negative imaginary part).[3] We are interested in the poles in the upper half plane, and the residue at these poles is given by

$$\mathrm{Res}(f, iG \pm H) = \lim_{u \to iG \pm H} (u - iG \mp H) \cdot f(u) \tag{C.44}$$

which, after some algebra, turns out to be

$$\mathrm{Res}(f, iG \pm H) = \frac{1}{i4H \cdot (\pm iG + H)} \tag{C.45}$$

So we have

$$\oint_C f(u)\,\mathrm{d}u = 2\pi i \cdot (\mathrm{Res}(f, iG + H) + \mathrm{Res}(f, iG - H)) \tag{C.46}$$

$$= 2\pi i \cdot \left(\frac{1}{i4H \cdot (iG + H)} + \frac{1}{i4H \cdot (-iG + H)} \right) \tag{C.47}$$

$$= \pi \cdot \frac{1}{G^2 + H^2} \tag{C.48}$$

$$= \pi \tag{C.49}$$

In the last step, we have used the fact that $G^2 + H^2 = 1$, since H is defined as $\sqrt{1 - G^2}$. Plugging this back into Eq. (C.40), we obtain

$$I = \frac{\pi \sqrt{\epsilon_s} L_p}{2c} \cdot \omega_P^2 \tag{C.50}$$

So we find that the integrated absorption is directly proportional to the squared plasma frequency ω_P^2. Notably, this expression does *not* depend on the other parameters in the dielectric tensor: the linewidth, γ, nor the centre frequency, ω_0. Thus – at least in the limit where the absorption coefficient is well-defined – the integrated absorption strength is directly tied to the plasma frequency of the QWs.

[3] To see why this is the case, recall that we defined $G = \gamma/2\omega_0$, which means that G is a positive, real number. If $G \leq 1$, then $H = \sqrt{1 - G^2}$ will be real, and it is clear that $iG \pm H$ will be in the upper half plane while $-iG \pm H$ will be in the lower half plane. If $G > 1$, we have $H = i\sqrt{G^2 - 1}$. Then, since $\sqrt{G^2 - 1} < G$, we will still have $iG \pm H = i\left(G \pm \sqrt{G^2 - 1}\right)$ in the upper half plane and $-iG \pm H = i\left(-G \pm \sqrt{G^2 - 1}\right)$ in the lower half plane. Thus, the statement holds for any $G > 0$.

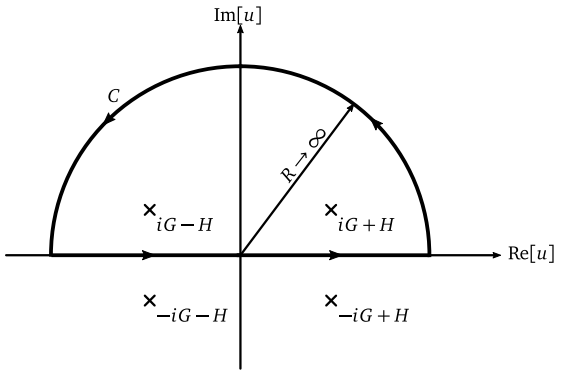

Figure C.2: The contour, C, used for residue integration of $f(u)$: a semicircle in the complex plane with the radius (R) taken to infinity. Also shown (×'s) are the poles of $f(u)$.

Appendix D

List of publications

- C. Deimert and Z. R. Wasilewski, "Precise control of time-varying effusion cell flux in molecular beam epitaxy", Journal of Vacuum Science & Technology A **39**, 043407 (2021)
- A. Khalatpour, A. K. Paulsen, C. Deimert, Z. R. Wasilewski, and Q. Hu, "High-power portable terahertz laser systems", Nature Photonics **15**, 16–20 (2021)
- C. Deimert, P. Goulain, J. M. Manceau, W. Pasek, T. Yoon, A. Bousseksou, N. Y. Kim, R. Colombelli, and Z. R. Wasilewski, "Realization of harmonic oscillator arrays with graded semiconductor quantum wells", Physical Review Letters **125**, 097403 (2020)
- B. Wen, C. Deimert, S. Wang, C. Xu, S. S. Rassel, Z. Wasilewski, and D. Ban, "Six-level hybrid extraction/injection scheme terahertz quantum cascade laser with suppressed thermally activated carrier leakage", Optics Express **28**, 26499–26508 (2020)
- C. Deimert and Z. R. Wasilewski, "MBE growth of continuously-graded parabolic quantum well arrays in AlGaAs", Journal of Crystal Growth **514**, 103–108 (2019)
- C. R. H. McRae, J. H. Béjanin, C. T. Earnest, T. G. McConkey, J. R. Rinehart, C. Deimert, J. P. Thomas, Z. R. Wasilewski, and M. Mariantoni, "Thin film metrology and microwave loss characterization of indium and aluminum/indium superconducting planar resonators", Journal of Applied Physics **123**, 205304 (2018)

Glossary

k · p The **k · p** method is an approach for simulating semiconductor band structures. See [76] for an introduction. xi, xiv, 35, 38, 39, 43, 45, 86, 88, 93, 100, 101

nextnano++ A commercial nanostructure simulator. https://www.nextnano.com/ x–xii, xiv, 34–36, 38, 39, 81, 86, 93